

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁵ : C12N 1/24, 15/00, C07H 21/00	A1	(11) International Publication Number: WO 92/06176 (43) International Publication Date: 16 April 1992 (16.04.92)
(21) International Application Number: PCT/US91/07141 (22) International Filing Date: 27 September 1991 (27.09.91) (30) Priority data: 590,664 28 September 1990 (28.09.90) US (71) Applicant: IXSYS, INC. [US/US]; 3550 General Atomics Court, Suite L103, San Diego, CA 92121 (US). (72) Inventor: HUSE, William, D. ; 471 Avenida Primavera, Del Mar, CA 92014 (US). (74) Agents: CAMPBELL, Cathryn et al.; Pretty, Schroeder, Brueggemann & Clark, 444 South Flower Street, Suite 2000, Los Angeles, CA 90071 (US).	(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC, MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, SD, SE (European patent), SN (OAPI patent), SU*, TD (OAPI patent), TG (OAPI patent). Published <i>With international search report.</i>	

(54) Title: SURFACE EXPRESSION LIBRARIES OF RANDOMIZED PEPTIDES**(57) Abstract**

A composition of matter comprising a plurality of procaryotic cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, said expressible oligonucleotides having a desirable bias of random codon sequences.

+ DESIGNATIONS OF "SU"

Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU+	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TC	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

SURFACE EXPRESSION LIBRARIES
OF RANDOMIZED PEPTIDES

5

BACKGROUND OF THE INVENTION

This invention relates generally to methods for synthesizing and expressing oligonucleotides and, more particularly, to methods for expressing oligonucleotides having random codon sequences.

10 Oligonucleotide synthesis proceeds via linear coupling of individual monomers in a stepwise reaction. The reactions are generally performed on a solid phase support by first coupling the 3' end of the first monomer to the support. The second monomer is added to the 5' end of the
15 first monomer in a condensation reaction to yield a dinucleotide coupled to the solid support. At the end of each coupling reaction, the by-products and unreacted, free monomers are washed away so that the starting material for the next round of synthesis is the pure oligonucleotide
20 attached to the support. In this reaction scheme, the stepwise addition of individual monomers to a single, growing end of a oligonucleotide ensures accurate synthesis of the desired sequence. Moreover, unwanted side reactions are eliminated, such as the condensation of two
25 oligonucleotides, resulting in high product yields.

In some instances, it is desired that synthetic oligonucleotides have random nucleotide sequences. This result can be accomplished by adding equal proportions of all four nucleotides in the monomer coupling reactions,
30 leading to the random incorporation of all nucleotides and yielding a population of oligonucleotides with random sequences. Since all possible combinations of nucleotide sequences are represented within the population, all possible codon triplets will also be represented. If the

objective is ultimately to generate random peptide products, this approach has a severe limitation because the random codons synthesized will bias the amino acids incorporated during translation of the DNA by the cell into
5 polypeptides.

The bias is due to the redundancy of the genetic code. There are four nucleotide monomers which leads to sixty-four possible triplet codons. With only twenty amino acids to specify, many of the amino acids are encoded by multiple
10 codons. Therefore, a population of oligonucleotides synthesized by sequential addition of monomers from a random population will not encode peptides whose amino acid sequence represents all possible combinations of the twenty different amino acids in equal proportions. That is, the
15 frequency of amino acids incorporated into polypeptides will be biased toward those amino acids which are specified by multiple codons.

To alleviate amino acid bias due to the redundancy of the genetic code, the oligonucleotides can be synthesized
20 from nucleotide triplets. Here, a triplet coding for each of the twenty amino acids is synthesized from individual monomers. Once synthesized, the triplets are used in the coupling reactions instead of individual monomers. By mixing equal proportions of the triplets, synthesis of
25 oligonucleotides with random codons can be accomplished. However, the cost of synthesis from such triplets far exceeds that of synthesis from individual monomers because triplets are not commercially available.

Amino acid bias can be reduced, however, by
30 synthesizing the degenerate codon sequence NNK where N is a mixture of all four nucleotides and K is a mixture of guanine and thymine nucleotides. Each position within an oligonucleotide having this codon sequence will contain a total of 32 codons (12 encoding amino acids being

represented once, 5 represented twice, 3 represented three times and one codon being a stop codon). Oligonucleotides expressed with such degenerate codon sequences will produce peptide products whose sequences are biased toward those amino acids being represented more than once. Thus, populations of peptides whose sequences are completely random cannot be obtained from oligonucleotides synthesized from degenerate sequences.

There thus exists a need for a method to express oligonucleotides having a fully random or desirably biased sequence which alleviates genetic redundancy. The present invention satisfies these needs and provides additional advantages as well.

SUMMARY OF THE INVENTION

The invention provides a plurality of procaryotic cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, the expressible oligonucleotides having a desirable bias of random codon sequences.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic drawing for synthesizing oligonucleotides from nucleotide monomers with random tuplets at each position using twenty reaction vessels.

Figure 2 is a schematic drawing for synthesizing oligonucleotides from nucleotide monomers with random tuplets at each position using ten reaction vessels.

Figure 3 is a schematic diagram of the two vectors used for sublibrary and library production from precursor oligonucleotide portions. M13IX22 (Figure 3A) is the vector used to clone the anti-sense precursor portions

(hatched box). The single-headed arrow represents the Lac p/o expression sequences and the double-headed arrow represents the portion of M13IX22 which is to be combined with M13IX42. The amber stop codon for biological selection and relevant restriction sites are also shown. 5 M13IX42 (Figure 3B) is the vector used to clone the sense precursor portions (open box). Thick lines represent the pseudo-wild type (ψ gVIII) and wild type (gVIII) gene VIII sequences. The double-headed arrow represents the portion 10 of M13IX42 which is to be combined with M13IX22. The two amber stop codons and relevant restriction sites are also shown. Figure 3C shows the joining of vector population from sublibraries to form the functional surface expression vector M13IX. Figure 3D shows the generation of a surface 15 expression library in a non-suppressor strain and the production of phage. The phage are used to infect a suppressor strain (Figure 3E) for surface expression and screening of the library.

Figure 4 is a schematic diagram of the vector used for 20 generation of surface expression libraries from random oligonucleotide populations (M13IX30). The symbols are as described for Figure 3.

Figure 5 is the nucleotide sequence of M13IX42 (SEQ ID NO: 1).

25 Figure 6 is the nucleotide sequence of M13IX22 (SEQ ID NO: 2).

Figure 7 is the nucleotide sequence of M13IX30 (SEQ ID NO: 3).

30 Figure 8 is the nucleotide sequence of M13ED03 (SEQ ID NO: 4).

Figure 9 is the nucleotide sequence of M13IX421 (SEQ

ID NO: 5).

Figure 10 is the nucleotide sequence of M13ED04 (SEQ ID NO: 6).

DETAILED DESCRIPTION OF THE INVENTION

5 This invention is directed to a simple and inexpensive method for synthesizing and expressing oligonucleotides having a desirable bias of random codons using individual monomers. The method is advantageous in that individual monomers are used instead of triplets and by synthesizing
10 only a non-degenerate subset of all triplets, codon redundancy is alleviated. Thus, the oligonucleotides synthesized represent a large proportion of possible random triplet sequences which can be obtained. The oligonucleotides can be expressed, for example, on the
15 surface of filamentous bacteriophage in a form which does not alter phage viability or impose biological selections against certain peptide sequences. The oligonucleotides produced are therefore useful for generating an unlimited number of pharmacological and research products.

20 In one embodiment, the invention entails the sequential coupling of monomers to produce oligonucleotides with a desirable bias of random codons. The coupling reactions for the randomization of twenty codons which specify the amino acids of the genetic code are performed
25 in ten different reaction vessels. Each reaction vessel contains a support on which the monomers for two different codons are coupled in three sequential reactions. One of the reactions couples an equal mixture of two monomers such that the final product has two different codon sequences.
30 The codons are randomized by removing the supports from the reaction vessels and mixing them to produce a single batch of supports containing all twenty codons at a particular position. Synthesis at the next codon position proceeds by

equally dividing the mixed batch of supports into ten reaction vessels as before and sequentially coupling the monomers for each pair of codons. The supports are again mixed to randomize the codons at the position just synthesized. The cycle of coupling, mixing and dividing continues until the desired number of codon positions have been randomized. After the last position has been randomized, the oligonucleotides with random codons are cleaved from the support. The random oligonucleotides can then be expressed, for example, on the surface of filamentous bacteriophage as gene VIII-peptide fusion proteins. Alternative genes can be used as well.

In its broadest form, the invention provides a diverse population of synthetic oligonucleotides contained in vectors so as to be expressible in cells. Such populations of diverse oligonucleotides can be fully random at one or more codon sites or can be fully defined at one or more site, so long as at least one site the codons are randomly variable. The populations of oligonucleotides can be expressed as fusion products in combination with surface proteins of filamentous bacteriophage, such as M13, as with gene VIII. The vectors can be transfected into a plurality of cells, such as the procaryote E. coli.

The diverse population of oligonucleotides can be formed by randomly combining first and second precursor populations, each precursor population having a desirable bias of random codon sequences. Methods of synthesizing and expressing the diverse population of expressible oligonucleotides are also provided.

In a preferred embodiment, two populations of random oligonucleotides are synthesized. The oligonucleotides within each population encode a portion of the final oligonucleotide which is to be expressed. Oligonucleotides within one population encode the carboxy terminal portion

of the expressed oligonucleotides. These oligonucleotides are cloned in frame with a gene VIII (gVIII) sequence so that translation of the sequence produces peptide fusion proteins. The second population of oligonucleotides are
5 cloned into a separate vector. Each oligonucleotide within this population encodes the anti-sense of the amino terminal portion of the expressed oligonucleotides. This vector also contains the elements necessary for expression. The two vectors containing the random oligonucleotides are
10 combined such that the two precursor oligonucleotide portions are joined together at random to form a population of larger oligonucleotides derived from two smaller portions. The vectors contain selectable markers to ensure maximum efficiency in joining together the two
15 oligonucleotide populations. A mechanism also exists to control the expression of gVIII-peptide fusion proteins during library construction and screening.

As used herein, the term "monomer" or "nucleotide monomer" refers to individual nucleotides used in the
20 chemical synthesis of oligonucleotides. Monomers that can be used include both the ribo- and deoxyribo- forms of each of the five standard nucleotides (derived from the bases adenine (A or dA, respectively), guanine (G or dG), cytosine (C or dC), thymine (T) and uracil (U)).
25 Derivatives and precursors of bases such as inosine which are capable of supporting polypeptide biosynthesis are also included as monomers. Also included are chemically modified nucleotides, for example, one having a reversible blocking agent attached to any of the positions on the
30 purine or pyrimidine bases, the ribose or deoxyribose sugar or the phosphate or hydroxyl moieties of the monomer. Such blocking groups include, for example, dimethoxytrityl, benzoyl, isobutyryl, beta-cyano thyl and diisopropylamine groups, and are used to protect hydroxyls, exocyclic amines
35 and phosphate moieties. Other blocking agents can also be used and are known to one skilled in the art.

As used herein, the term "tuplet" refers to a group of elements of a definable size. The elements of a tuplet as used herein are nucleotide monomers. For example, a tuplet can be a dinucleotide, a trinucleotide or can also be four or more nucleotides.

As used herein, the term "codon" or "triplet" refers to a tuplet consisting of three adjacent nucleotide monomers which specify one of the twenty naturally occurring amino acids found in polypeptide biosynthesis. The term also includes nonsense, or stop, codons which do not specify any amino acid.

"Random codons" or "randomized codons," as used herein, refers to more than one codon at a position within a collection of oligonucleotides. The number of different codons can be from two to twenty at any particular position. "Randomized oligonucleotides," as used herein, refers to a collection of oligonucleotides with random codons at one or more positions. "Random codon sequences" as used herein means that more than one codon position within a randomized oligonucleotide contains random codons. For example, if randomized oligonucleotides are six nucleotides in length (i.e., two codons) and both the first and second codon positions are randomized to encode all twenty amino acids, then a population of oligonucleotides having random codon sequences with every possible combination of the twenty triplets in the first and second position makes up the above population of randomized oligonucleotides. The number of possible codon combinations is 20^2 . Likewise, if randomized oligonucleotides of fifteen nucleotides in length are synthesized which have random codon sequences at all positions encoding all twenty amino acids, then all triplets coding for each of the twenty amino acids will be found in equal proportions at every position. The population constituting the randomized oligonucleotides

will contain 20^{15} different possible species of oligonucleotides. "Random tuplets," or "randomized tuplets" are defined analogously.

As used herein, the term "bias" refers to a preference. It is understood that there can be degrees of preference or bias toward codon sequences which encode particular amino acids. For example, an oligonucleotide whose codon sequences do not preferably encode particular amino acids is unbiased and therefore completely random. The oligonucleotide codon sequences can also be biased toward predetermined codon sequences or codon frequencies and while still diverse and random, will exhibit codon sequences biased toward a defined, or preferred, sequence. "A desirable bias of random codon sequences" as used herein, refers to the predetermined degree of bias which can be selected from totally random to essentially, but not totally, defined (or preferred). There must be at least one codon position which is variable, however.

As used herein, the term "support" refers to a solid phase material for attaching monomers for chemical synthesis. Such support is usually composed of materials such as beads of control pore glass but can be other materials known to one skilled in the art. The term is also meant to include one or more monomers coupled to the support for additional oligonucleotide synthesis reactions.

As used herein, the terms "coupling" or "condensing" refers to the chemical reactions for attaching one monomer to a second monomer or to a solid support. Such reactions are known to one skilled in the art and are typically performed on an automated DNA synthesizer such as a MilliGen/Bioscience Cyclone Plus Synthesizer using procedures recommended by the manufacturer. "Sequentially coupling" as used herein, refers to the stepwise addition of monomers.

A method of synthesizing oligonucleotides having random tuplets using individual monomers is described. The method consists of several steps, the first being synthesis of a nucleotide tuplet for each tuplet to be randomized. As described here and below, a nucleotide triplet (i.e., a codon) will be used as a specific example of a tuplet. Any size tuplet will work using the methods disclosed herein, and one skilled in the art would know how to use the methods to randomize tuplets of any size.

10 If the randomization of codons specifying all twenty amino acids is desired at a position, then twenty different codons are synthesized. Likewise, if randomization of only ten codons at a particular position is desired then those ten codons are synthesized. Randomization of codons from 15 two to sixty-four can be accomplished by synthesizing each desired triplet. Preferably, randomization of from two to twenty codons is used for any one position because of the redundancy of the genetic code. The codons selected at one position do not have to be the same codons selected at the 20 next position. Additionally, the sense or anti-sense sequence oligonucleotide can be synthesized. The process therefore provides for randomization of any desired codon position with any number of codons.

Codons to be randomized are synthesized sequentially 25 by coupling the first monomer of each codon to separate supports. The supports for the synthesis of each codon can, for example, be contained in different reaction vessels such that one reaction vessel corresponds to the monomer coupling reactions for one codon. As will be used 30 here and below, if twenty codons are to be randomized, then twenty reaction vessels can be used in independent coupling reactions for the first twenty monomers of each codon. Synthesis proceeds by sequentially coupling the second monomer of each codon to the first monomer to produce a 35 dimer, followed by coupling the third monomer for each

codon to each of the above-synthesized dimers to produce a trimer (Figure 1, step 1, where M_1 , M_2 and M_3 represent the first, second and third monomer, respectively, for each codon to be randomized).

5 Following synthesis of the first codons from individual monomers, the randomization is achieved by mixing the supports from all twenty reaction vessels which contain the individual codons to be randomized. The solid phase support can be removed from its vessel and mixed to
10 achieve a random distribution of all codon species within the population (Figure 1, step 2). The mixed population of supports, constituting all codon species, are then redistributed into twenty independent reaction vessels
15 (Figure 1, step 3). The resultant vessels are all identical and contain equal portions of all twenty codons coupled to a solid phase support.

For randomization of the second position codon, synthesis of twenty additional codons is performed in each of the twenty reaction vessels produced in step 3 as the
20 condensing substrates of step 1 (Figure 1, step 4). Steps 1 and 4 are therefore equivalent except that step 4 uses the supports produced by the previous synthesis cycle (steps 1 through 3) for codon synthesis whereas step 1 is the initial synthesis of the first codon in the
25 oligonucleotide. The supports resulting from step 4 will each have two codons attached to them (i.e., a hexanucleotide) with the codon at the first position being any one of twenty possible codons (i.e., random) and the codon at the second position being one of the twenty
30 possible codons.

For randomization of the codon at the second position and synthesis of the third position codon, steps 2 through 4 are again repeated. This process yields in each vessel a three codon oligonucleotide (i.e., 9 nucleotides) with

codon positions 1 and 2 randomized and position three containing one of the twenty possible codons. Steps 2 through 4 are repeated to randomize the third position codon and synthesize the codon at the next position. The process is continued until an oligonucleotide of the desired length is achieved. After the final randomization step, the oligonucleotide can be cleaved from the supports and isolated by methods known to one skilled in the art. Alternatively, the oligonucleotides can remain on the supports for use in methods employing probe hybridization.

The diversity of codon sequences, i.e., the number of different possible oligonucleotides, which can be obtained using the methods of the present invention, is extremely large and only limited by the physical characteristics of available materials. For example, a support composed of beads of about 100 μm in diameter will be limited to about 10,000 beads/reaction vessel using a 1 μM reaction vessel containing 25 mg of beads. This size bead can support about 1×10^7 oligonucleotides per bead. Synthesis using separate reaction vessels for each of the twenty amino acids will produce beads in which all the oligonucleotides attached to an individual bead are identical. The diversity which can be obtained under these conditions is approximately 10^7 copies of 10,000 x 20 or 200,000 different random oligonucleotides. The diversity can be increased, however, in several ways without departing from the basic methods disclosed herein. For example, the number of possible sequences can be increased by decreasing the size of the individual beads which make up the support. A bead of about 30 μm in diameter will increase the number of beads per reaction vessel and therefore the number of oligonucleotides synthesized. Another way to increase the diversity of oligonucleotides with random codons is to increase the volume of the reaction vessel. For example, using the same size bead, a larger volume can contain a greater number of beads than a smaller vessel and therefore

support the synthesis of a greater number of oligonucleotides. Increasing the number of codons coupled to a support in a single reaction vessel also increases the diversity of the random oligonucleotides. The total
5 diversity will be the number of codons coupled per vessel raised to the number of codon positions synthesized. For example, using ten reaction vessels, each synthesizing two codons to randomize a total of twenty codons, the number of different oligonucleotides of ten codons in length per 100
10 μm bead can be increased where each bead will contain about 2^{10} or 1×10^3 different sequences instead of one. One skilled in the art will know how to modify such parameters to increase the diversity of oligonucleotides with random codons.

15 A method of synthesizing oligonucleotides having random codons at each position using individual monomers wherein the number of reaction vessels is less than the number of codons to be randomized is also described. For
20 example, if twenty codons are to be randomized at each position within an oligonucleotide population, then ten reaction vessels can be used. The use of a smaller number of reaction vessels than the number of codons to be randomized at each position is preferred because the
25 and results in a greater number of possible oligonucleotides synthesized.

The use of a smaller number of reaction vessels for random synthesis of twenty codons at a desired position within an oligonucleotide is similar to that described
30 above using twenty reaction vessels except that each reaction vessel can contain the synthesis products of more than one codon. For example, step on synthesis using ten reaction vessels proceeds by coupling about two different codons on supports contained in each of ten reaction
35 vessels. This is shown in Figure 2 where each of the two

codons coupled to a different support can consist of the following sequences: (1) (T/G)TT for Phe and Val; (2) (T/C)CT for Ser and Pro; (3) (T/C)AT for Tyr and His; (4) (T/C)GT for Cys and Arg; (5) (C/A)TG for Leu and Met; (6) (C/G)AG for Gln and Glu; (7) (A/G)CT for Thr and Ala; (8) (A/G)AT for Asn and Asp; (9) (T/G)GG for Trp and Gly and (10) A(T/A)A for Ile and Cys. The slash (/) signifies that a mixture of the monomers indicated on each side of the slash are used as if they were a single monomer in the indicated coupling step. The antisense sequence for each of the above codons can be generated by synthesizing the complementary sequence. For example, the antisense for Phe and Val can be AA(C/A). The amino acids encoded by each of the above pairs of sequences are given as the standard three letter nomenclature.

Coupling of the monomers in this fashion will yield codons specifying all twenty of the naturally occurring amino acids attached to supports in ten reaction vessels. However, the number of individual reaction vessels to be used will depend on the number of codons to be randomized at the desired position and can be determined by one skilled in the art. For example, if ten codons are to be randomized, then five reaction vessels can be used for coupling. The codon sequences given above can be used for this synthesis as well. The sequences of the codons can also be changed to incorporate or be replaced by any of the additional forty-four codons which constitutes the genetic code.

The remaining steps of synthesis of oligonucleotides with random codons using a smaller number of reaction vessels are as outlined above for synthesis with twenty reaction vessels except that the mixing and dividing steps are performed with supports from about half the number of reaction vessels. These remaining steps are shown in Figure 2 (steps 2 through 4).

Oligonucleotides having at least one specified tuplet at a predetermined position and the remaining positions having random tuplets can also be synthesized using the methods described herein. The synthesis steps are similar to those outlined above using twenty or less reaction vessels except that prior to synthesis of the specified codon position, the dividing of the supports into separate reaction vessels for synthesis of different codons is omitted. For example, if the codon at the second position of the oligonucleotide is to be specified, then following synthesis of random codons at the first position and mixing of the supports, the mixed supports are not divided into new reaction vessels but, instead, can be contained in a single reaction vessel to synthesize the specified codon. The specified codon is synthesized sequentially from individual monomers as described above. Thus, the number of reaction vessels can be increased or decreased at each step to allow for the synthesis of a specified codon or a desired number of random codons.

Following codon synthesis, the mixed supports are divided into individual reaction vessels for synthesis of the next codon to be randomized (Figure 1, step 3) or can be used without separation for synthesis of a consecutive specified codon. The rounds of synthesis can be repeated for each codon to be added until the desired number of positions with predetermined or randomized codons are obtained.

Synthesis of oligonucleotides with the first position codon being specified can also be synthesized using the above method. In this case, the first position codon is synthesized from the appropriate monomers. The supports are divided into the required number of reaction vessels needed for synthesis of random codons at the second position and the rounds of synthesis, mixing and dividing are performed as described above.

A method of synthesizing oligonucleotides having
tuplets which are diverse but biased toward a predetermined
sequence is also described herein. This method employs two
reaction vessels, one vessel for the synthesis of a
5 predetermined sequence and the second vessel for the
synthesis of a random sequence. This method is
advantageous to use when a significant number of codon
positions, for example, are to be of a specified sequence
since it alleviates the use of multiple reaction vessels.
10 Instead, a mixture of four different monomers such as
adenine, guanine, cytosine and thymine nucleotides are used
for the first and second monomers in the codon. The codon
is completed by coupling a mixture of a pair of monomers of
either guanine and thymine or cytosine and adenine
15 nucleotides at the third monomer position. In the second
vessel, nucleotide monomers are coupled sequentially to
yield the predetermined codon sequence. Mixing of the two
supports yields a population of oligonucleotides containing
both the predetermined codon and the random codons at the
20 desired position. Synthesis can proceed by using this
mixture of supports in a single reaction vessel, for
example, for coupling additional predetermined codons or,
further dividing the mixture into two reaction vessels for
synthesis of additional random codons.

25 The two reaction vessel method can be used for codon
synthesis within an oligonucleotide with a predetermined
tuplet sequence by dividing the support mixture into two
portions at the desired codon position to be randomized.
Additionally, this method allows for the extent of
30 randomization to be adjusted. For example, unequal mixing
or dividing of the two supports will change the fraction of
codons with predetermined sequences compared to those with
random codons at the desired position. Unequal mixing and
dividing of supports can be useful when there is a need to
35 synthesize random codons at a significant number of
positions within an oligonucleotide of a longer or shorter

length.

The extent of randomization can also be adjusted by using unequal mixtures of monomers in the first, second and third monomer coupling steps of the random codon position.

5 The unequal mixtures can be in any or all of the coupling steps to yield a population of codons enriched in sequences reflective of the monomer proportions.

Synthesis of randomized oligonucleotides is performed using methods well known to one skilled in the art. Linear

10 coupling of monomers can, for example, be accomplished using phosphoramidite chemistry with a MilliGen/Biosearch Cyclone Plus automated synthesizer as described by the manufacturer (Millipore, Burlington, MA). Other chemistries and automated synthesizers can be employed as

15 well and are known to one skilled in the art.

Synthesis of multiple codons can be performed without modification to the synthesizer by separately synthesizing the codons in individual sets of reactions. Alternatively, modification of an automated DNA synthesizer can be

20 performed for the simultaneous synthesis of codons in multiple reaction vessels.

In one embodiment, the invention provides a plurality of procaryotic cells containing a diverse population of expressible oligonucleotides operationally linked to

25 expression elements, the expressible oligonucleotides having a desirable bias of random codon sequences produced from diverse combinations of first and second oligonucleotides having a desirable bias of random sequences. The invention provides for a method for

30 constructing such a plurality of procaryotic cells as well.

The oligonucleotides synthesized by the above methods can be used to express a plurality of random peptides which

are unbiased, diverse but biased toward a predetermined sequence or which contain at least one specified codon at a predetermined position. The need will determine which type of oligonucleotide is to be expressed to give the resultant population of random peptides and is known to one skilled in the art. Expression can be performed in any compatible vector/host system. Such systems include, for example, plasmids or phagemids in procaryotes such as *E. coli*, yeast systems, and other eucaryotic systems such as mammalian cells, but will be described herein in context with its presently preferred embodiment, i.e. expression on the surface of filamentous bacteriophage. Filamentous bacteriophage can be, for example, M13, f1 and fd. Such phage have circular single-stranded genomes and double strand replicative DNA forms. Additionally, the peptides can also be expressed in soluble or secreted form depending on the need and the vector/host system employed.

Expression of random peptides on the surface of M13 can be accomplished, for example, using the vector system shown in Figure 3. Construction of the vectors enabling one of ordinary skill to make them are explicitly set out in Examples I and II. The complete nucleotide sequences are given in Figures 5, 6 and 7 (SEQ ID NOS: 1, 2 and 3, respectively). This system produces random oligonucleotides functionally linked to expression elements and to gVIII by combining two smaller oligonucleotide portions contained in separate vectors into a single vector. The diversity of oligonucleotide species obtained by this system or others described herein can be 5×10^7 or greater. Diversity of less than 5×10^7 can also be obtained and will be determined by the need and type of random peptides to be expressed. The random combination of two precursor portions into a larger oligonucleotide increases the diversity of the population several fold and has the added advantage of producing oligonucleotides larger than what can be synthesized by standard methods.

Additionally, although the correlation is not known, when the number of possible paths an oligonucleotide can take during synthesis such as described herein is greater than the number of beads, then there will be a correlation between the synthesis path and the sequences obtained. By combining oligonucleotide populations which are synthesized separately, this correlation will be destroyed. Therefore, any bias which may be inherent in the synthesis procedures will be alleviated by joining two precursor portions into a contiguous random oligonucleotide.

Populations of precursor oligonucleotides to be combined into an expressible form are each cloned into separate vectors. The two precursor portions which make up the combined oligonucleotide corresponds to the carboxy and amino terminal portions of the expressed peptide. Each precursor oligonucleotide can encode either the sense or anti-sense and will depend on the orientation of the expression elements and the gene encoding the fusion portion of the protein as well as the mechanism used to join the two precursor oligonucleotides. For the vectors shown in Figure 3, precursor oligonucleotides corresponding to the carboxy terminal portion of the peptide encode the sense strand. Those corresponding to the amino terminal portion encode the anti-sense strand. Oligonucleotide populations are inserted between the Eco RI and Sac I restriction enzyme sites in M13IX22 and M13IX42 (Figure 3A and B). M13IX42 (SEQ ID NO: 1) is the vector used for sense strand precursor oligonucleotide portions and M13IX22 (SEQ ID NO: 2) is used for anti-sense precursor portions.

The populations of randomized oligonucleotides inserted into the vectors are synthesized with Eco RI and Sac I recognition sequences flanking opposite ends of the random codon sequences. The sites allow annealing and ligation of the single strand oligonucleotides into a double stranded vector restricted with Eco RI and Sac I.

Alternatively, the oligonucleotides can be inserted into the vector by standard mutagenesis methods. In this latter method, single stranded vector DNA is isolated from the phage and annealed with random oligonucleotides having
5 known sequences complementary to vector sequences. The oligonucleotides are extended with DNA polymerase to produce double stranded vectors containing the randomized oligonucleotides.

The vector used for sense strand oligonucleotide
10 portions, M13IX42 (Figure 3B) contains down-stream and in frame with the Eco RI and Sac I restriction sites a sequence encoding the pseudo-wild type gVIII product. This gene encodes the wild type M13 gVIII amino acid sequence but has been changed at the nucleotide level to reduce
15 homologous recombination with the wild type gVIII contained on the same vector. The wild type gVIII is present to ensure that at least some functional, non-fusion coat protein will be produced. The inclusion of a wild type gVIII therefore reduces the possibility of non-viable phage
20 production and biological selection against certain peptide fusion proteins. Differential regulation of the two genes can also be used to control the relative ratio of the pseudo and wild type proteins.

Also contained downstream and in frame with the Eco RI
25 and Sac I restriction sites is an amber stop codon. The mutation is located six codons downstream from Sac I and therefore lies between the inserted oligonucleotides and the gVIII sequence. As was the function of the wild type gVIII, the amber stop codon also reduces biological
30 selection when combining precursor portions to produce expressible oligonucleotides. This is accomplished by using a non-suppressor (sup 0) host strain because non-suppressor strains will terminate expression after the olig nucleotide sequences but before the pseudo gVIII
35 sequences. Therefore, the pseudo gVIII will never be

expressed on the phage surface under these circumstances. Instead, only soluble peptides will be produced. Expression in a non-suppressor strain can be advantageously utilized when one wishes to produce large populations of soluble peptides. Stop codons other than amber, such as opal and ochre, or molecular switches, such as inducible repressor elements, can also be used to unlink peptide expression from surface expression. Additional controls exist as well and are described below.

10 The vector used for anti-sense strand oligonucleotide portions, M13IX22, (Figure 3A), contains the expression elements for the peptide fusion proteins. Upstream and in frame with the Sac I and Eco RI sites in this vector is a leader sequence for surface expression. A ribosome binding site and Lac Z promoter/operator elements are present for transcription and translation of the peptide fusion proteins.

Both vectors contain a pair of Fok I restriction enzyme sites (Figure 3 A and B) for joining together two precursor oligonucleotide portions and their vector sequences. One site is located at the ends of each precursor oligonucleotide which is to be joined. The second Fok I site within the vectors is located at the end of the vector sequences which are to be joined. The 5' overhang of this second Fok I site has been altered to encode a sequence which is not found in the overhangs produced at the first Fok I site within the oligonucleotide portions. The two sites allow the cleavage of each circular vector into two portions and subsequent ligation of essential components within each vector into a single circular vector where the two oligonucleotide precursor portions form a contiguous sequence (Figure 3C). Non-compatible overhangs produced at the two Fok I sites allows optimal conditions to be selected for performing concatermization or circularization reactions for joining

the two vector portions. Such selection of conditions can be used to govern the reaction order and therefore increase the efficiency of joining.

Fok I is a restriction enzyme whose recognition sequence is distal to the point of cleavage. Distal placement of the recognition sequence in its location to the cleavage point is important since if the two were superimposed within the oligonucleotide portions to be combined, it would lead to an invariant codon sequence at the juncture. To alleviate the formation of invariant codons at the juncture, Fok I recognition sequences can be placed outside of the random codon sequence and still be used to restrict within the random sequence. Subsequent annealing of the single-strand overhangs produced by Fok I and ligation of the two oligonucleotide precursor portions allows the juncture to be formed. A variety of restriction enzymes restrict DNA by this mechanism and can be used instead of Fok I to join precursor oligonucleotides without creating invariant codon sequences. Such enzymes include, for example, Alw I, Bbu I, Bsp MI, Hga I, Hph I, Mbo II, Mnl I, Ple I and Sfa NI. One skilled in the art knows how to substitute Fok I recognition sequences for alternative enzyme recognition sequences such as those above, and use the appropriate enzyme for joining precursor oligonucleotide portions.

Although the sequences of the precursor oligonucleotides are random and will invariably have oligonucleotides within the two precursor populations whose sequences are sufficiently complementary to anneal after cleavage, the efficiency of annealing can be increased by insuring that the single-strand overhangs within one precursor population will have a complementary sequence within the second precursor population. This can be accomplished by synthesizing a non-degenerate series of known sequences at the Fok I cleavage site coding for each

of the twenty amino acids. Since the Fok I cleavage site contains a four base overhang, forty different sequences are needed to randomly encode all twenty amino acids. For example, if two precursor populations of ten codons in length are to be combined, then after the ninth codon position is synthesized, the mixed population of supports are divided into forty reaction vessels for each of the populations and complementary sequences for each of the corresponding reaction vessels between populations are independently synthesized. The sequences are shown in Tables III and VI of Example I where the oligonucleotides on columns 1R through 40R form complementary overhangs with the oligonucleotides on the corresponding columns 1L through 40L once cleaved. The degenerate X positions in Table VI are necessary to maintain the reading frame once the precursor oligonucleotide portions are joined. However, use of restriction enzymes which produce a blunt end, such as Mnl I can be alternatively used in place of Fok I to alleviate the degeneracy introduced in maintaining the reading frame.

The last feature exhibited by each of the vectors is an amber stop codon located in an essential coding sequence within the vector portion lost during combining (Figure 3C). The amber stop codon is present to select for viable phage produced from only the proper combination of precursor oligonucleotides and their vector sequences into a single vector species. Other non-sense mutations or selectable markers can work as well.

The combining step randomly brings together different precursor oligonucleotides within the two populations into a single vector (Figure 3C; M13IX). The vector sequences donated from each independent vector, M13IX22 and M13IX42, are necessary for production of viable phage. Also, since the expression elements are contained in M13IX22 and the gVIII sequences are contained in M13IX42, expression of

functional gVIII-peptide fusion proteins cannot be accomplished until the sequences are linked as shown in M13IX.

5 The combining step is performed by restricting each population of vectors containing randomized oligonucleotides with Fok I, mixing and ligating (Figure 3C). Any vectors generated which contain an amber stop codon will not produce viable phage when introduced into a non-suppressor strain (Figure 3D). Therefore, only the sequences which do not contain an amber stop codon will make up the final population of vectors contained in the library. These vector sequences are the sequences required for surface expression of randomized peptides. By analogous methodology, more than two vector portions can be combined into a single vector which expresses random peptides.

20 The invention provides for a method of selecting peptides capable of being bound by a ligand binding protein from a population of random peptides by (a) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector; (b) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors; (d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of random peptides; and (e) determining the peptides which bind to said binding protein. The invention also provides for determining the encoding nucleic acid sequence of such peptides as well.

Surface expression of the random peptide library is performed in an amber suppressor strain. As described above, the amber stop codon between the random codon sequence and the gVIII sequence unlinks the two components in a non-suppressor strain. Isolating the phage produced from the non-suppressor strain and infecting a suppressor strain will link the random codon sequences to the gVIII sequence during expression (Figure 3E). Culturing the suppressor strain after infection allows the expression of all peptide species within the library as gVIII-peptide fusion proteins. Alternatively, the DNA can be isolated from the non-suppressor strain and then introduced into a suppressor strain to accomplish the same effect.

The level of expression of gVIII-peptide fusion proteins can additionally be controlled at the transcriptional level. The gVIII-peptide fusion proteins are under the inducible control of the Lac Z promoter/operator system. Other inducible promoters can work as well and are known by one skilled in the art. For high levels of surface expression, the suppressor library is cultured in an inducer of the Lac Z promoter such as isopropylthio- β -galactoside (IPTG). Inducible control is beneficial because biological selection against non-functional gVIII-peptide fusion proteins can be minimized by culturing the library under non-expressing conditions. Expression can then be induced only at the time of screening to ensure that the entire population of oligonucleotides within the library are accurately represented on the phage surface. Also this can be used to control the valency of the peptide on the phage surface.

The surface expression library is screened for specific peptides which bind ligand binding proteins by standard affinity isolation procedures. Such methods include, for example, panning, affinity chromatography and solid phase blotting procedures. Panning as described by

Parmley and Smith, Gene 73:305-318 (1988), which is incorporated herein by reference, is preferred because high titers of phage can be screened easily, quickly and in small volumes. Furthermore, this procedure can select
5 minor peptide species within the population, which otherwise would have been undetectable, and amplified to substantially homogenous populations. The selected peptide sequences can be determined by sequencing the nucleic acid encoding such peptides after amplification of the phage
10 population.

The invention provides a plurality of procaryotic cells containing a diverse population of oligonucleotides having a desirable bias of random codon sequences that are operationally linked to expression sequences. The
15 invention provides for methods of constructing such populations of cells as well.

Random oligonucleotides synthesized by any of the methods described previously can also be expressed on the surface of filamentous bacteriophage, such as M13, for
20 example, without the joining together of precursor oligonucleotides. A vector such as that shown in Figure 4, M13IX30, can be used. This vector exhibits all the functional features of the combined vector shown in Figure 3C for surface expression of gVIII-peptide fusion proteins.
25 The complete nucleotide sequence for M13IX30 (SEQ ID NO: 3) is shown in Figure 7.

M13IX30 contains a wild type gVIII for phage viability and a pseudo gVIII sequence for peptide fusions. The vector also contains in frame restriction sites for cloning
30 random peptides. The cloning sites in this vector are Xho I, Stu I and Spe I. Oligonucleotides should therefore be synthesized with the appropriate complementary ends for annealing and ligation or insertional mutagenesis. Alternatively, the appropriate termini can be generated by

PCR technology. Between the restriction sites and the pseudo gVIII sequence is an in-frame amber stop codon, again, ensuring complete viability of phage in constructing and manipulating the library. Expression and screening is performed as described above for the surface expression library of oligonucleotides generated from precursor portions.

Thus, the invention provides a method of selecting peptides capable of being bound by a ligand binding protein from a population of random peptides by (a) operationally linking a diverse population of oligonucleotides having a desirable bias of random codon sequences to expression elements; (b) introducing said population of vectors into a compatible host under conditions sufficient for expressing said population of random peptides; and (c) determining the peptides which bind to said binding protein. Also provided is a method for determining the encoding nucleic acid sequence of such selected peptides.

The following examples are intended to illustrate, but not limit the invention.

EXAMPLE I

Isolation and Characterization of Peptide Ligands Generated From Right and Left Half Random Oligonucleotides

This example shows the synthesis of random oligonucleotides and the construction and expression of surface expression libraries of the encoded randomized peptides. The random peptides of this example derive from the mixing and joining together of two random oligonucleotides. Also demonstrated is the isolation and characterization of peptide ligands and their corresponding nucleotide sequence for specific binding proteins.

Synthesis of Random Oligonucleotides

The synthesis of two randomized oligonucleotides which correspond to smaller portions of a larger randomized oligonucleotide is shown below. Each of the two smaller portions make up one-half of the larger oligonucleotide. The population of randomized oligonucleotides constituting each half are designated the right and left half. Each population of right and left halves are ten codons in length with twenty random codons at each position. The right half corresponds to the sense sequence of the randomized oligonucleotides and encode the carboxy terminal half of the expressed peptides. The left half corresponds to the anti-sense sequence of the randomized oligonucleotides and encode the amino terminal half of the expressed peptides. The right and left halves of the randomized oligonucleotide populations are cloned into separate vector species and then mixed and joined so that the right and left halves come together in random combination to produce a single expression vector species which contains a population of randomized oligonucleotides twenty codons in length. Electroporation of the vector population into an appropriate host produces filamentous phage which express the random peptides on their surface.

The reaction vessels for oligonucleotide synthesis were obtained from the manufacturer of the automated synthesizer (Millipore, Burlington, MA; supplier of MilliGen/Biosearch Cyclone Plus Synthesizer). The vessels were supplied as packages containing empty reaction columns (1 μ mole), frits, crimps and plugs (MilliGen/Biosearch catalog # GEN 860458). Derivatized and underivatized control pore glass, phosphoramidite nucleotides, and synthesis reagents were also obtained from MilliGen/Biosearch. Crimper and decrimper tools were obtained from Fisher Scientific Co., Pittsburgh, PA (Catalog numbers 06-406-20 and 06-406-25A, respectively).

Ten reaction columns were used for right half synthesis of random oligonucleotides ten codons in length. The oligonucleotides have 5 monomers at their 3' end of the sequence 5'GAGCT3' and 8 monomers at their 5' end of the sequence 5'AATTCCAT3'. The synthesizer was fitted with a column derivatized with a thymine nucleotide (T-column, MilliGen/Biosearch # 0615.50) and was programmed to synthesize the sequences shown in Table I for each of ten columns in independent reaction sets. The sequence of the last three monomers (from right to left since synthesis proceeds 3' to 5') encode the indicated amino acids:

Table I

	<u>Column</u>	<u>Sequence (5' to 3')</u>	<u>Amino Acids</u>
15	column 1R	(T/G)TTGAGCT	Phe and Val
	column 2R	(T/C)CTGAGCT	Ser and Pro
	column 3R	(T/C)ATGAGCT	Tyr and His
	column 4R	(T/C)GTGAGCT	Cys and Arg
	column 5R	(C/A)TGGAGCT	Leu and Met
20	column 6R	(C/G)AGGAGCT	Gln and Glu
	column 7R	(A/G)CTGAGCT	Thr and Ala
	column 8R	(A/G)ATGAGCT	Asn and Asp
	column 9R	(T/G)GGGAGCT	Trp and Gly
	column 1R	A(T/A)AGAGCT	Ile and Cys

where the two monomers in parentheses denote a single monomer position within the codon and indicate that an equal mixture of each monomer was added to the reaction for coupling. The monomer coupling reactions for each of the 10 columns were performed as recommended by the manufacturer (amidit version S1.06, # 8400-050990, scale 1 μ M). After the last coupling reaction, the columns were washed with acetonitrile and lyophilized to dryness.

Following synthesis, the plugs were removed from each

column using a decrimper and the reaction products were poured into a single weigh boat. Initially the bead mass increases, due to the weight of the monomers, however, at later rounds of synthesis material is lost. In either case, the material was equalized with underivatized control pore glass and mixed thoroughly to obtain a random distribution of all twenty codon species. The reaction products were then aliquotted into 10 new reaction columns by removing 25 mg of material at a time and placing it into separate reaction columns. Alternatively, the reaction products can be aliquotted by suspending the beads in a liquid that is dense enough for the beads to remain dispersed, preferably a liquid that is equal in density to the beads, and then aliquoting equal volumes of the suspension into separate reaction columns. The lip on the inside of the columns where the frits rest was cleared of material using vacuum suction with a syringe and 25 G needle. New frits were placed onto the lips, the plugs were fitted into the columns and were crimped into place using a crimper.

Synthesis of the second codon position was achieved using the above 10 columns containing the random mixture of reaction products from the first codon synthesis. The monomer coupling reactions for the second codon position are shown in Table II. An A in the first position means that any monomer can be programmed into the synthesizer. At that position, the first monomer position is not coupled by the synthesizer since the software assumes that the monomer is already attached to the column. An A also denotes that the columns from the previous codon synthesis should be placed on the synthesizer for use in the present synthesis round. Reactions were again sequentially repeated for each column as shown in Table II and the reaction products washed and dried as described above.

Table II

	<u>Column</u>	<u>Sequence</u> <u>(5' to 3')</u>	<u>Amino Acids</u>
	column 1R	(T/G)TTA	Phe and Val
5	column 2R	(T/C)CTA	Ser and Pro
	column 3R	(T/C)ATA	Tyr and His
	column 4R	(T/C)GTA	Cys and Arg
	column 5R	(C/A)TGA	Leu and Met
	column 6R	(C/G)AGA	Gln and Glu
10	column 7R	(A/G)CTA	Thr and Ala
	column 8R	(A/G)ATA	Asn and Asp
	column 9R	(T/G)GGA	Trp and Gly
	column 10R	A(T/A)AA	Ile and Cys

Randomization of the second codon position was achieved by removing the reaction products from each of the columns and thoroughly mixing the material. The material was again divided into new reaction columns and prepared for monomer coupling reactions as described above.

Random synthesis of the next seven codons (positions 3 through 9) proceeded identically to the cycle described above for the second codon position and again used the monomer sequences of Table II. Each of the newly repacked columns containing the random mixture of reaction products from synthesis of the previous codon position was used for the synthesis of the subsequent codon position. After synthesis of the codon at position nine and mixing of the reaction products, the material was divided and repacked into 40 different columns and the monomer sequences shown in Table III were coupled to each of the 40 columns in independent reactions. The oligonucleotides from each of the 40 columns were mixed once more and cleaved from the control pore glass as recommended by the manufacturer.

Table III

	<u>Column</u>	<u>Sequence (5' to 3')</u>
	column 1R	AATTCTTTTA
5	column 2R	AATTCTGTTA
	column 3R	AATTCGTTA
	column 4R	AATTCGGTTA
	column 5R	AATTCTTCTA
	column 6R	AATTCTCCTA
10	column 7R	AATTCGTCTA
	column 8R	AATTCGCCTA
	column 9R	AATTCTTATA
	column 10R	AATTCTCATA
	column 11R	AATTCGTATA
15	column 12R	AATTCGCATA
	column 13R	AATTCTTGTA
	column 14R	AATTCTCGTA
	column 15R	AATTCGTGTA
	column 16R	AATTCGCGTA
20	column 17R	AATTCTCTGA
	column 18R	AATTCTATGA
	column 19R	AATTCGCTGA
	column 20R	AATTCGATGA
	column 21R	AATTCTCAGA
25	column 22R	AATTCTGAGA
	column 23R	AATTCGCAGA
	column 24R	AATTCGGAGA
	column 25R	AATTCTACTA
	column 26R	AATTCTGCTA
30	column 27R	AATTCGACTA
	column 28R	AATTCGGCTA
	column 29R	AATTCTAATA
	column 30R	AATTCTGATA
	column 31R	AATTCGAATA
35	column 32R	AATTCGGATA
	column 33R	AATTCTTGGA

	column 34R	AATTCTGGGA
	column 35R	AATTCGTGGA
	column 36R	AATTCGGGGA
	column 37R	AATTCTATAA
5	column 38R	AATTCTAAAA
	column 39R	AATTCGATAA
	column 40R	AATTCGAAAA

Left half synthesis of random oligonucleotides proceeded similarly to the right half synthesis. This half of the oligonucleotide corresponds to the anti-sense sequence of the encoded randomized peptides. Thus, the complementary sequence of the codons in Tables I through III are synthesized. The left half oligonucleotides also have 5 monomers at their 3' end of the sequence 5'GAGCT3' and 8 monomers at their 5' end of the sequence 5'AATTCCAT3'. The rounds of synthesis, washing, drying, mixing, and dividing are as described above.

For the first codon position, the synthesizer was fitted with a T-column and programmed to synthesize the sequences shown in Table IV for each of ten columns in independent reaction sets. As with right half synthesis, the sequence of the last three monomers (from right to left) encode the indicated amino acids:

Table IV

	<u>Column</u>	<u>Sequence (5' to 3')</u>	<u>Amino Acids</u>
	column 1L	AA(A/C)GAGCT	Phe and Val
5	column 2L	AG(A/G)GAGCT	Ser and Pro
	column 3L	AT(A/G)GAGCT	Tyr and His
	column 4L	AC(A/G)GAGCT	Cys and Arg
	column 5L	CA(G/T)GAGCT	Leu and Met
	column 6L	CT(G/C)GAGCT	Gln and Glu
10	column 7L	AG(T/C)GAGCT	Thr and Ala
	column 8L	AT(T/C)GAGCT	Asn and Asp
	column 9L	CC(A/C)GAGCT	Trp and Gly
	column 10L	T(A/T)TGAGCT	Ile and Cys

Following washing and drying, the plugs for each column were removed, mixed and aliquotted into ten new reaction columns as described above. Synthesis of the second codon position was achieved using these ten columns containing the random mixture of reaction products from the first codon synthesis. The monomer coupling reactions for the second codon position are shown in Table V.

Table V

	<u>Column</u>	<u>Sequence (5' to 3')</u>	<u>Amino Acids</u>
	column 1L	AA(A/C) <u>A</u>	Phe and Val
25	column 2L	AG(A/G) <u>A</u>	Ser and Pro
	column 3L	AT(A/G) <u>A</u>	Tyr and His
	column 4L	AC(A/G) <u>A</u>	Cys and Arg
	column 5L	CA(G/T) <u>A</u>	Leu and Met
	column 6L	CT(G/C) <u>A</u>	Gln and Glu
30	column 7L	AG(T/C) <u>A</u>	Thr and Ala
	column 8L	AT(T/C) <u>A</u>	Asn and Asp
	column 9L	CC(A/C) <u>A</u>	Trp and Gly
	column 10L	T(A/T) <u>TA</u>	Ile and Cys

Again, randomization of the second codon position was achieved by removing the reaction products from each of the columns and thoroughly mixing the beads. The beads were repacked into ten new reaction columns.

5 Random synthesis of the next seven codon positions proceeded identically to the cycle described above for the second codon position and again used the monomer sequences of Table V. After synthesis of the codon at position nine
 10 and mixing of the reaction products, the material was divided and repacked into 40 different columns and the monomer sequences shown in Table VI were coupled to each of the 40 columns in independent reactions.

Table VI

	<u>Column</u>	<u>Sequence (5' to 3')</u>
15	column 1L	AATTCCATAAAAXXA
	column 2L	AATTCCATAAACXXA
	column 3L	AATTCCATAACAXXA
	column 4L	AATTCCATAACCXXA
20	column 5L	AATTCCATAGAAXXA
	column 6L	AATTCCATAGACXXA
	column 7L	AATTCCATAGGAXXA
	column 8L	AATTCCATAGGCXXA
	column 9L	AATTCCATATAAXXA
25	column 10L	AATTCCATATACXXA
	column 11L	AATTCCATATGAXXA
	column 12L	AATTCCATATGCXXA
	column 13L	AATTCCATACAAXXA
	column 14L	AATTCCATACACXXA
30	column 15L	AATTCCATACGAXXA
	column 16L	AATTCCATACGCXXA
	column 17L	AATTCCATCAGAXXA
	column 18L	AATTCCATCAGCXXA
	column 19L	AATTCCATCATAXXA
35	column 20L	AATTCCATCATCXXA

	column 21L	AATTCCATCTGAXXA
	column 22L	AATTCCATCTGCXXA
	column 23L	AATTCCATCTCAXXA
	column 24L	AATTCCATCTCCXXA
5	column 25L	AATTCCATAGTAXXA
	column 26L	AATTCCATAGTCXXA
	column 27L	AATTCCATAGCAXXA
	column 28L	AATTCCATAGCCXXA
	column 29L	AATTCCATATTAXXA
10	column 30L	AATTCCATATTCXXA
	column 31L	AATTCCATATCAXXA
	column 32L	AATTCCATATCCXXA
	column 33L	AATTCCATCCAAXXA
	column 34L	AATTCCATCCACXXA
15	column 35L	AATTCCATCCCAXXA
	column 36L	AATTCCATCCCCXXA
	column 37L	AATTCCATTATAXXA
	column 38L	AATTCCATTATCXXA
	column 39L	AATTCCATTTTAXXA
20	column 40L	AATTCCATTTTCXXA

The first two monomers denoted by an "X" represent an equal mixture of all four nucleotides at that position. This is necessary to retain a relatively unbiased codon sequence at the junction between right and left half oligonucleotides.

25 The above right and left half random oligonucleotides were cleaved and purified from the supports and used in constructing the surface expression libraries below.

Vector Construction

Two M13-based vectors, M13IX42 (SEQ ID NO: 1) and

30 M13IX22 (SEQ ID NO: 2), were constructed for the cloning and propagation of right and left half populations of random oligonucleotides, respectively. The vectors were specially constructed to facilitate the random joining and subsequent expression of right and left half

oligonucleotide populations. Each vector within the population contains one right and one left half oligonucleotide from the population joined together to form a single contiguous oligonucleotide with random codons which is twenty-two codons in length. The resultant population of vectors are used to construct a surface expression library.

M13IX42, or the right-half vector, was constructed to harbor the right half populations of randomized oligonucleotides. M13mp18 (Pharmacia, Piscataway, NJ) was the starting vector. This vector was genetically modified to contain, in addition to the encoded wild type M13 gene VIII already present in the vector: (1) a pseudo-wild type M13 gene VIII sequence with a stop codon (amber) placed between it and an Eco RI-Sac I cloning site for randomized oligonucleotides; (2) a pair of Fok I sites to be used for joining with M13IX22, the left-half vector; (3) a second amber stop codon placed on the opposite side of the vector than the portion being combined with the left-half vector; and (4) various other mutations to remove redundant restriction sites and the amino terminal portion of Lac Z.

The pseudo-wild type M13 gene VIII was used for surface expression of random peptides. The pseudo-wild type gene encodes the identical amino acid sequence as that of the wild type gene; however, the nucleotide sequence has been altered so that only 63% identity exists between this gene and the encoded wild type gene VIII. Modification of the gene VIII nucleotide sequence used for surface expression reduces the possibility of homologous recombination with the wild type gene VIII contained on the same vector. Additionally, the wild type M13 gene VIII was retained in the vector system to ensure that at least some functional, non-fusion coat protein would be produced. The inclusion of wild type gene VIII therefore reduces the possibility of non-viable phage production from the random

peptide fusion genes.

The pseudo-wild type gene VIII was constructed by chemically synthesizing a series of oligonucleotides which encode both strands of the gene. The oligonucleotides are presented in Table VII (SEQ ID NOS: 7 through 16).

TABLE VII

Pseudo-Wild Type Gene VIII Oligonucleotide Series

	<u>Top Strand Oligonucleotides</u>	<u>Sequence (5' to 3')</u>
10	VIII 03	GATCC TAG GCT GAA GGC GAT
		GAC CCT GCT AAG GCT GC
	VIII 04	A TTC AAT AGT TTA CAG GCA
		AGT GCT ACT GAG TAC A
	VIII 05	TT GGC TAC GCT TGG GCT ATG
15		GTA GTA GTT ATA GTT
	VIII 06	GGT GCT ACC ATA GGG ATT AAA
		TTA TTC AAA AAG TT
	VIII 07	T ACG AGC AAG GCT TCT TA
20	<u>Bottom Strand Oligonucleotides</u>	
	VIII 08	AGC TTA AGA AGC CTT GCT CGT
		AAA CTT TTT GAA TAA TTT
	VIII 09	AAT CCC TAT GGT AGC ACC AAC
		TAT AAC TAC TAC CAT
25	VIII 10	AGC CCA AGC GTA GCC AAT GTA
		CTC AGT AGC ACT TG
	VIII 11	C CTG TAA ACT ATT GAA TGC
		AGC CTT AGC AGG GTC
	VIII 12	ATC GCC TTC AGC CTA G

30 Except for the terminal oligonucleotides VIII 03 (SEQ

ID NO: 7) and VIII 08 (SEQ ID NO: 12), the above oligonucleotides (oligonucleotides VIII 04-VIII 07 and 09-12 (SEQ ID NOS: 8 through 11 and 13 through 16)) were mixed at 200 ng each in 10 μ l final volume and phosphorylated with T4 polynucleotide Kinase (Pharmacia, Piscataway, NJ) with 1 mM ATP at 37°C for 1 hour. The reaction was stopped at 65°C for 5 minutes. Terminal oligonucleotides were added to the mixture and annealed into double-stranded form by heating to 65°C for 5 minutes, followed by cooling to room temperature over a period of 30 minutes. The annealed oligonucleotides were ligated together with 1.0 U of T4 DNA ligase (BRL). The annealed and ligated oligonucleotides yield a double-stranded DNA flanked by a Bam HI site at its 5' end and by a Hind III site at its 3' end. A translational stop codon (amber) immediately follows the Bam HI site. The gene VIII sequence begins with the codon GAA (Glu) two codons 3' to the stop codon. The double-stranded insert was phosphorylated using T4 DNA Kinase (Pharmacia, Piscataway, NJ) and ATP (10 mM Tris-HCl, pH 7.5, 10 mM MgCl₂) and cloned in frame with the Eco RI and Sac I sites within the M13 polylinker. To do so, M13mp18 was digested with Bam HI (New England Biolabs, Beverly, MA) and Hind III (New England Biolabs) and combined at a molar ratio of 1:10 with the double-stranded insert. The ligations were performed at 16°C overnight in 1X ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM MgCl₂, 20 mM DTT, 1 mM ATP, 50 μ g/ml BSA) containing 1.0 U of T4 DNA ligase (New England Biolabs). The ligation mixture was transformed into a host and screened for positive clones using standard procedures in the art.

Several mutations were generated within the right-half vector to yield functional M13IX42. The mutations were generated using the method of Kunkel et al., Meth. Enzymol. 154:367-382 (1987), which is incorporated herein by reference, for site-directed mutagenesis. The reagents, strains and protocols were obtained from a Bio Rad

Mutagenesis kit (Bio Rad, Richmond, CA) and mutagenesis was performed as recommended by the manufacturer.

A Fok I site used for joining the right and left halves was generated 8 nucleotides 5' to the unique Eco RI site using the oligonucleotide 5'-CTCGAATTCGTACATCCTGGTCATAGC-3' (SEQ ID NO: 17). The second Fok I site retained in the vector is naturally encoded at position 3547; however, the sequence within the overhang was changed to encode CTTC. Two Fok I sites were removed from the vector at positions 239 and 7244 of M13mp18 as well as the Hind III site at the end of the pseudo gene VIII sequence using the mutant oligonucleotides 5'-CATTTTTCAGATGGCTTAGA-3' (SEQ ID NO: 18) and 5'-TAGCATTAACGTCCAATA-3' (SEQ ID NO: 19), respectively. New Hind III and Mlu I sites were also introduced at position 3919 and 3951 of M13IX42. The oligonucleotides used for this mutagenesis had the sequences 5'-ATATATTTTAGTAAGCTTCATCTTCT-3' (SEQ ID NO: 20) and 5'-GACAAAGAACGCGTGAAACTTT-3' (SEQ ID NO: 21), respectively. The amino terminal portion of Lac Z was deleted by oligonucleotide-directed mutagenesis using the mutant oligonucleotide 5'-GCGGGCCTCTTCGCTATTGCTTAAGAAGCCTTGCT-3' (SEQ ID NO: 22). This deletion also removed a third M13mp18 derived Fok I site. The distance between the Eco RI and Sac I sites was increased to ensure complete double digestion by inserting a spacer sequence. The spacer sequence was inserted using the oligonucleotide 5'-TTCAGCCTAGGATCCGCCGAGCTCTCCTACCTGCGAATTCGTACATCC-3' (SEQ ID NO: 23). Finally, an amber stop codon was placed at position 4492 using the mutant oligonucleotide 5'-TGGATTATACTTCTA AATAATGGA-3' (SEQ ID NO: 24). The amber stop codon is used as a biological selection to ensure the proper recombination of vector sequences to bring together right and left halves of the randomized oligonucleotides. In constructing the above mutations, all changes made in a M13 coding region were performed such that the amino acid

sequence remained unaltered. It should be noted that several mutations within M13mp18 were found which differed from the published sequence. Where known, these sequence differences are recorded herein as found and therefore may not correspond exactly to the published sequence of M13mp18.

The sequence of the resultant vector, M13IX42, is shown in Figure 5 (SEQ ID NO: 1). Figure 3A also shows M13IX42 where each of the elements necessary for producing a surface expression library between right and left half randomized oligonucleotides is marked. The sequence between the two Fok I sites shown by the arrow is the portion of M13IX42 which is to be combined with a portion of the left-half vector to produce random oligonucleotides as fusion proteins of gene VIII.

M13IX22, or the left-half vector, was constructed to harbor the left half populations of randomized oligonucleotides. This vector was constructed from M13mp19 (Pharmacia, Piscataway, NJ) and contains: (1) Two Fok I sites for mixing with M13IX42 to bring together the left and right halves of the randomized oligonucleotides; (2) sequences necessary for expression such as a promoter and signal sequence and translation initiation signals; (3) an Eco RI-Sac I cloning site for the randomized oligonucleotides; and (4) an amber stop codon for biological selection in bringing together right and left half oligonucleotides.

Of the two Fok I sites used for mixing M13IX22 with M13IX42, one is naturally encoded in M13mp18 and M13mp19 (at position 3547). As with M13IX42, the overhang within this naturally occurring Fok I site was changed to CTTC. The other Fok I site was introduced after construction of the translation initiation signals by site-directed mutagenesis using the oligonucleotide 5'-

TAACACTCATTCCGGATGGAATTCTGGAGTCTGGGT-3' (SEQ ID NO: 25).

The translation initiation signals were constructed by annealing of overlapping oligonucleotides as described above to produce a double-stranded insert containing a 5' Eco RI site and a 3' Hind III site. The overlapping oligonucleotides are shown in Table VIII (SEQ ID NOS: 26 through 34) and were ligated as a double-stranded insert between the Eco RI and Hind III sites of M13mp18 as described for the pseudo gene VIII insert. The ribosome binding site (AGGAGAC) is located in oligonucleotide 015 (SEQ ID NO: 26) and the translation initiation codon (ATG) is the first three nucleotides of oligonucleotide 016 (SEQ ID NO: 27).

TABLE VIII

Oligonucleotide Series for Construction of
Translation Signals in M13IX22

<u>Oligonucleotide</u>	<u>Sequence (5' to 3')</u>
015	AATT C GCC AAG GAG ACA GTC AT
016	AATG AAA TAC CTA TTG CCT ACG GCA
017	GCC GCT GGA TTG TT
018	ATTA CTC GCT GCC CAA CCA GCC ATG
019	GCC GAG CTC GTG AT
020	GACC CAG ACT CCA GATATC CAA CAG
021	GAA TGA GTG TTA AT
022	TCT AGA ACG CGT C
023	ACGT G ACG CGT TCT AGA AT TAA
024	CACTCA TTC CTG T
025	TG GAT ATC TGG AGT CTG GGT CAT
026	CAC GAG CTC GGC CAT G
027	GC TGG TTG GGC AGC GAG TAA TAA
028	CAA TCC AGC GGC TGC C
029	GT AGG CAA TAG GTA TTT CAT TAT
030	GAC TGT CCT TGG CG

Oligonucleotide 017 (SEQ ID NO: 27) contained a Sac I restriction site 67 nucleotides downstream from the ATG codon. The naturally occurring Eco RI site was removed and a new site introduced 25 nucleotides downstream from the
5 Sac I. Oligonucleotides 5'-TGACTGTCTCCTTGGCGTGTGAAATTGTTA-3' (SEQ ID NO: 35) and 5'-TAACACTCATTCCGGATGGAATTCTGGAGTCTGGGT-3' (SEQ ID NO: 36) were used to generate each of the mutations, respectively. An amber stop codon was also introduced at position 3263 of M13mp18 using the
10 oligonucleotide 5'-CAATTTTATCCTAAATCTTACCAAC-3' (SEQ ID NO: 37).

In addition to the above mutations, a variety of other modifications were made to remove certain sequences and redundant restriction sites. The LAC Z ribosome binding
15 site was removed when the original Eco RI site in M13mp18 was mutated. Also, the Fok I sites at positions 239, 6361 and 7244 of M13mp18 were likewise removed with mutant oligonucleotides 5'-CATTTTGCAGATGGCTTAGA-3' (SEQ ID NO: 38), 5'-CGAAAGGGGGGTGTGCTGCAA-3' (SEQ ID NO: 39) and 5'-
20 TAGCATTAACGTCCAATA-3' (SEQ ID NO: 40), respectively. Again, mutations within the coding region did not alter the amino acid sequence.

The resultant vector, M13IX22, is 7320 base pairs in length, the sequence of which is shown in Figure 6 (SEQ ID
25 NO: 2). The Sac I and Eco RI cloning sites are at positions 6290 and 6314, respectively. Figure 3A also shows M13IX22 where each of the elements necessary for producing a surface expression library between right and left half randomized oligonucleotides is marked.

30 Library Construction

Each population of right and left half randomized oligonucleotides from columns 1R through 40R and columns 1L through 40L are cloned separately into M13IX42 and M13IX22,

respectively, to create sublibraries of right and left half randomized oligonucleotides. Therefore, a total of eighty sublibraries are generated. Separately maintaining each population of randomized oligonucleotides until the final
5 screening step is performed to ensure maximum efficiency of annealing of right and left half oligonucleotides. The greater efficiency increases the total number of randomized oligonucleotides which can be obtained. Alternatively, one can combine all forty populations of right half
10 oligonucleotides (columns 1R-40R) into one population and of left half oligonucleotides (columns 1L-40L) into a second population to generate just one sublibrary for each.

For the generation of sublibraries, each of the above populations of randomized oligonucleotides are cloned
15 separately into the appropriate vector. The right half oligonucleotides are cloned into M13IX42 to generate sublibraries M13IX42.1R through M13IX42.40R. The left half oligonucleotides are similarly cloned into M13IX22 to generate sublibraries M13IX22.1L through M13IX22.40L. Each
20 vector contains unique Eco RI and Sac I restriction enzyme sites which produce 5' and 3' single-stranded overhangs, respectively, when digested. The single strand overhangs are used for the annealing and ligation of the complementary single-stranded random oligonucleotides.

25 The randomized oligonucleotide populations are cloned between the Eco RI and Sac I sites by sequential digestion and ligation steps. Each vector is treated with an excess of Eco RI (New England Biolabs) at 37°C for 2 hours followed by addition of 4-24 units of calf intestinal
30 alkaline phosphatase (Boehringer Mannheim, Indianapolis, IN). Reactions are stopped by phenol/chloroform extraction and ethanol precipitation. The pellets are resuspended in an appropriate amount of distilled or deionized water (dH₂O). About 10 pmol of vector is mixed with a 5000-fold
35 molar excess of each population of randomized

oligonucleotides in 10 μ l of 1X ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, 50 μ g/ml BSA) containing 1.0 U of T4 DNA ligase (BRL, Gaithersburg, MD). The ligation is incubated at 16°C for 16 hours. Reactions
5 are stopped by heating at 75°C for 15 minutes and the DNA is digested with an excess of Sac I (New England Biolabs) for 2 hours. Sac I is inactivated by heating at 75°C for 15 minutes and the volume of the reaction mixture is adjusted to 300 μ l with an appropriate amount of 10X ligase
10 buffer and dH_2O . One unit of T4 DNA ligase (BRL) is added and the mixture is incubated overnight at 16°C. The DNA is ethanol precipitated and resuspended in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). DNA from each ligation is electroporated into XL1 BlueTM cells (Stratagene, La Jolla,
15 CA), as described below, to generate the sublibraries.

E. coli XL1 BlueTM is electroporated as described by Smith et al., Focus 12:38-40 (1990) which is incorporated herein by reference. The cells are prepared by inoculating a fresh colony of XL1s into 5 mls of SOB without magnesium
20 (20 g bacto-tryptone, 5 g bacto-yeast extract, 0.584 g NaCl, 0.186 g KCl, dH_2O to 1,000 mls) and grown with vigorous aeration overnight at 37°C. SOB without magnesium (500 ml) is inoculated at 1:1000 with the overnight culture and grown with vigorous aeration at 37°C until the OD_{550} is
25 0.8 (about 2 to 3 h). The cells are harvested by centrifugation at 5,000 rpm (2,600 x g) in a GS3 rotor (Sorvall, Newtown, CT) at 4°C for 10 minutes, resuspended in 500 ml of ice-cold 10% (v/v) sterile glycerol and centrifuged and resuspended a second time in the same
30 manner. After a third centrifugation, the cells are resuspended in 10% sterile glycerol at a final volume of about 2 ml, such that the OD_{550} of the suspension is 200 to 300. Usually, resuspension is achieved in the 10% glycerol that remains in the bottle after pouring off the supernate.
35 Cells are frozen in 40 μ l aliquots in microcentrifuge tubes using a dry ice-ethanol bath and stored frozen at -70°C.

Frozen cells are electroporated by thawing slowly on ice before use and mixing with about 10 pg to 500 ng of vector per 40 μ l of cell suspension. A 40 μ l aliquot is placed in an 0.1 cm electroporation chamber (Bio-Rad, Richmond, CA) and pulsed once at 0°C using 200 Ω parallel resistor, 25 μ F, 1.88-kV, which gives a pulse length (τ) of 4 ms. A 10 μ l aliquot of the pulsed cells are diluted into 1 ml SOC (98 mls SOB plus 1 ml of 2 M $MgCl_2$ and 1 ml of 2 M glucose) in a 12- x 75-mm culture tube, and the culture is shaken at 37°C for 1 hour prior to culturing in selective media, (see below).

Each of the eighty sublibraries are cultured using methods known to one skilled in the art. Such methods can be found in Sanbrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1989, and in Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, New York, 1989, both of which are incorporated herein by reference. Briefly, the above 1 ml sublibrary cultures were grown up by diluting 50-fold into 2XYT media (16 g tryptone, 10 g yeast extract, 5 g NaCl) and culturing at 37°C for 5-8 hours. The bacteria were pelleted by centrifugation at 10,000 xg. The supernatant containing phage was transferred to a sterile tube and stored at 4°C.

Double strand vector DNA containing right and left half randomized oligonucleotide inserts is isolated from the cell pellet of each sublibrary. Briefly, the pellet is washed in TE (10 mM Tris, pH 8.0, 1 mM EDTA) and recollected by centrifugation at 7,000 rpm for 5' in a Sorval centrifuge (Newtown, CT). Pellets are resuspended in 6 mls of 10% Sucrose, 50 mM Tris, pH 8.0. 3.0 ml of 10 mg/ μ l lysozyme is added and incubated on ice for 20 minutes. 12 mls of 0.2 M NaOH, 1% SDS is added followed by 10 minutes on ice. The suspensions are then incubated on ice for 20 minutes after addition of 7.5 mls of 3 M NaOAc,

pH 4.6. The samples are centrifuged at 15,000 rpm for 15 minutes at 4°C, RNased and extracted with phenol/chloroform, followed by ethanol precipitation. The pellets are resuspended, weighed and an equal weight of CsCl₂ is dissolved into each tube until a density of 1.60 g/ml is achieved. EtBr is added to 600 µg/ml and the double-stranded DNA is isolated by equilibrium centrifugation in a TV-1665 rotor (Sorval) at 50,000 rpm for 6 hours. These DNAs from each right and left half sublibrary are used to generate forty libraries in which the right and left halves of the randomized oligonucleotides have been randomly joined together.

Each of the forty libraries are produced by joining together one right half and one left half sublibrary. The two sublibraries joined together corresponded to the same column number for right and left half random oligonucleotide synthesis. For example, sublibrary M13IX42.1R is joined with M13IX22.1L to produce the surface expression library M13IX.1RL. In the alternative situation where only two sublibraries are generated from the combined populations of all right half synthesis and all left half synthesis, only one surface expression library would be produced.

For the random joining of each right and left half oligonucleotide populations into a single surface expression vector species, the DNAs isolated from each sublibrary are digested an excess of Fok I (New England Biolabs). The reactions are stopped by phenol/chloroform extraction, followed by ethanol precipitation. Pellets are resuspended in dH₂O. Each surface expression library is generated by ligating equal molar amounts (5-10 pmol) of Fok I digested DNA isolated from corresponding right and left half sublibraries in 10 µl of 1X ligase buffer containing 1.0 U of T4 DNA ligase (Bethesda Research Laboratories, Gaithersburg, MD). The ligations proceed

overnight at 16°C and are electroporated into the sup O strain MK30-3 (Boehringer Mannheim Biochemical, (BMB), Indianapolis, IN) as previously described for XL1 cells. Because MK30-3 is sup O, only the vector portions encoding the randomized oligonucleotides which come together will produce viable phage.

Screening of Surface Expression Libraries

Purified phage are prepared from 50 ml liquid cultures of XL1 BlueTM cells (Stratagene) which are infected at a m.o.i. of 10 from the phage stocks stored at 4°C. The cultures are induced with 2 mM IPTG. Supernatants from all cultures are combined and cleared by two centrifugations, and the phage are precipitated by adding 1/7.5 volumes of PEG solution (25% PEG-8000, 2.5 M NaCl), followed by incubation at 4°C overnight. The precipitate is recovered by centrifugation for 90 minutes at 10,000 x g. Phage pellets are resuspended in 25 ml of 0.01 M Tris-HCl, pH 7.6, 1.0 mM EDTA, and 0.1% Sarkosyl and then shaken slowly at room temperature for 30 minutes. The solutions are adjusted to 0.5 M NaCl and to a final concentration of 5% polyethylene glycol. After 2 hours at 4°C, the precipitates containing the phage are recovered by centrifugation for 1 hour at 15,000 X g. The precipitates are resuspended in 10 ml of NET buffer (0.1 M NaCl, 1.0 mM EDTA, and 0.01 M Tris-HCl, pH 7.6), mixed well, and the phage repelleted by centrifugation at 170,000 X g for 3 hours. The phage pellets are subsequently resuspended overnight in 2 ml of NET buffer and subjected to cesium chloride centrifugation for 18 hours at 110,000 X g (3.86 g of cesium chloride in 10 ml of buffer). Phage bands are collected, diluted 7-fold with NET buffer, recentrifuged at 170,000 X g for 3 hours, resuspended, and stored at 4°C in 0.3 ml of NET buffer containing 0.1 mM sodium azide.

Ligand binding proteins used for panning on

streptavidin coated dishes are first biotinylated and then absorbed against UV-inactivated blocking phage (see below). The biotinylating reagents are dissolved in dimethylformamide at a ratio of 2.4 mg solid NHS-SS-Biotin (sulfo-succinimidyl 2-(biotinamido)ethyl-1,3'-dithiopropionate; Pierce, Rockford, IL) to 1 ml solvent and used as recommended by the manufacturer. Small-scale reactions are accomplished by mixing 1 μ l dissolved reagent with 43 μ l of 1 mg/ml ligand binding protein diluted in sterile bicarbonate buffer (0.1 M NaHCO_3 , pH 8.6). After 2 hours at 25°C, residual biotinylating reagent is reacted with 500 μ l 1 M ethanolamine (pH adjusted to 9 with HCl) for an additional 2 hours. The entire sample is diluted with 1 ml TBS containing 1 mg/ml BSA, concentrated to about 50 μ l on a Centricon 30 ultra-filter (Amicon), and washed on the same filter three times with 2 ml TBS and once with 1 ml TBS containing 0.02% NaN_3 and 7×10^{12} UV-inactivated blocking phage (see below); the final retentate (60-80 μ l) is stored at 4°C. Ligand binding proteins biotinylated with the NHS-SS-Biotin reagent are linked to biotin via a disulfide-containing chain.

UV-irradiated M13 phage were used for blocking binding proteins which fortuitously bound filamentous phage in general. M13mp8 (Messing and Vieira, Gene 19: 262-276 (1982), which is incorporated herein by reference) was chosen because it carries two amber stop codons, which ensure that the few phage surviving irradiation will not grow in the sup O strains used to titer the surface expression libraries. A 5 ml sample containing 5×10^{13} M13mp8 phage, purified as described above, was placed in a small petri plate and irradiated with a germicidal lamp at a distance of two feet for 7 minutes (flux 150 $\mu\text{W}/\text{cm}^2$). NaN_3 was added to 0.02% and phage particles concentrated to 10^{14} particles/ml on a Centricon 30-kDa ultrafilter (Amicon).

For panning, polystyrene petri plates (60 x 15 mm, Falcon; Becton Dickinson, Lincoln Park, NJ) are incubated with 1 ml of 1 mg/ml of streptavidin (BMB) in 0.1 M NaHCO₃ pH 8.6-0.02% NaN₃ in a small, air-tight plastic box overnight in a cold room. The next day streptavidin is removed and replaced with at least 10 ml blocking solution (29 mg/ml of BSA; 3 µg/ml of streptavidin; 0.1 M NaHCO₃ pH 8.6-0.02% NaN₃) and incubated at least 1 hour at room temperature. The blocking solution is removed and plates are washed rapidly three times with Tris buffered saline containing 0.5% Tween 20 (TBS-0.5% Tween 20).

Selection of phage expressing peptides bound by the ligand binding proteins is performed with 5 µl (2.7 µg ligand binding protein) of blocked biotinylated ligand binding proteins reacted with a 50 µl portion of each library. Each mixture is incubated overnight at 4°C, diluted with 1 ml TBS-0.5% Tween 20, and transferred to a streptavidin-coated petri plate prepared as described above. After rocking 10 minutes at room temperature, unbound phage are removed and plates washed ten times with TBS-0.5% Tween 20 over a period of 30-90 minutes. Bound phage are eluted from plates with 800 µl sterile elution buffer (1 mg/ml BSA, 0.1 M HCl, pH adjusted to 2.2 with glycerol) for 15 minutes and eluates neutralized with 48 µl 2 M Tris (pH unadjusted). A 20 µl portion of each eluate is titered on MK30-3 concentrated cells with dilutions of input phage.

A second round of panning is performed by treating 750 µl of first eluate from each library with 5 mM DTT for 10 minutes to break disulfide bonds linking biotin groups to residual biotinylated binding proteins. The treated eluate is concentrated on a Centricon 30 ultrafilter (Amicon), washed three times with TBS-0.5% Tween 20, and concentrated to a final volume of about 50 µl. Final retentate is transferred to a tube containing 5.0 µl (2.7 µg ligand

binding protein) blocked biotinylated ligand binding proteins and incubated overnight. The solution is diluted with 1 ml TBS-0.5% Tween 20, panned, and eluted as described above on fresh streptavidin-coated petri plates.

5 The entire second eluate (800 μ l) is neutralized with 48 μ l 2 M Tris, and 20 μ l is titered simultaneously with the first eluate and dilutions of the input phage.

Individual phage populations are purified through 2 to 3 rounds of plaque purification. Briefly, the second

10 eluate titer plates are lifted with nitrocellulose filters (Schleicher & Schuell, Inc., Keene, NH) and processed by washing for 15 minutes in TBS (10 mM Tris-HCl, pH 7.2, 150 mM NaCl), followed by an incubation with shaking for an additional 1 hour at 37°C with TBS containing 5% nonfat dry

15 milk (TBS-5% NDM) at 0.5 ml/cm². The wash is discarded and fresh TBS-5% NDM is added (0.1 ml/cm²) containing the ligand binding protein between 1 nM to 100 mM, preferably between 1 to 100 μ M. All incubations are carried out in heat-sealable pouches (Sears). Incubation with the ligand

20 binding protein proceeds for 12-16 hours at 4°C with shaking. The filters are removed from the bags and washed 3 times for 30 minutes at room temperature with 150 mls of TBS containing 0.1% NDM and 0.2% NP-40 (Sigma, St. Louis, MO). The filters are then incubated for 2 hours at room

25 temperature in antiserum against the ligand binding protein at an appropriate dilution in TBS-0.5% NDM, washed in 3 changes of TBS containing 0.1% NDM and 0.2% NP-40 as described above and incubated in TBS containing 0.1% NDM and 0.2% NP-40 with 1×10^6 cpm of ¹²⁵I-labeled Protein A

30 (specific activity = 2.1×10^7 cpm/ μ g). After a washing with TBS containing 0.1% NDM and 0.2% NP-40 as described above, the filters are wrapped in Saran Wrap and exposed to Kodak X-Omat x-ray film (Kodak, Rochester, NY) for 1-12 hours at -70°C using Dupont Cronex Lightning Plus

35 Intensifying Screens (Dupont, Willmington, DE).

Positive plaques identified are cored with the large end of a pasteur pipet and placed into 1 ml of SM (5.8 g NaCl, 2 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 50 ml 1 M Tris-HCl, pH 7.5, 5 mls 2% gelatin, to 1000 mls with dH_2O) plus 1-3 drops of CHCl_3 , and
5 incubated at 37°C 2-3 hours or overnight at 4°C. The phage are diluted 1:500 in SM and 2 μl are added to 300 μl of XL1 cells plus 3 mls of soft agar per 100 mm^2 plate. The XL1 cells are prepared for plating by growing a colony overnight in 10 ml LB (10 g bacto-tryptone, 5 g bacto-yeast
10 extract, 10 g NaCl, 1000 ml dH_2O) containing 100 μl of 20% maltose and 100 μl of 1 M MgSO_4 . The bacteria are pelleted by centrifugation at 2000 xg for 10 minutes and the pellet is resuspended gently in 10 mls of 10 mM MgSO_4 . The
15 suspension is diluted 4-fold by adding 30 mls of 10 mM MgSO_4 to give an OD_{600} of approximately 0.5. The second and third round screens are identical to that described above except that the plaques are cored with the small end of a pasteur pipet and placed into 0.5 mls SM plus a drop of CHCl_3 , and 1-
5 μl of the phage following incubation are used for plating
20 without dilution. At the end of the third round of purification, an individual plaque is picked and the templates prepared for sequencing.

Template Preparation and Sequencing

25 Templates are prepared for sequencing by inoculating a 1 ml culture of 2XYT containing a 1:100 dilution of an overnight culture of XL1 with an individual plaque. The plaques are picked using a sterile toothpick. The culture is incubated at 37°C for 5-6 hours with shaking and then
30 transferred to a 1.5 ml microfuge tube. 200 μl of PEG solution is added, followed by vortexing and placed on ice for 10 minutes. The phage precipitate is recovered by centrifugation in a microfuge at 12,000 x g for 5 minutes. The supernatant is discarded and the pellet is resuspended
35 in 230 μl of TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) by gently pipeting with a yellow pipet tip. Phenol (200 μl)

is added, followed by a brief vortex and microfuged to separate the phases. The aqueous phase is transferred to a separate tube and extracted with 200 μ l of phenol/chloroform (1:1) as described above for the phenol extraction. A 0.1 volume of 3 M NaOAc is added, followed by addition of 2.5 volumes of ethanol and precipitated at -20°C for 20 minutes. The precipitated templates are recovered by centrifugation in a microfuge at 12,000 x g for 8 minutes. The pellet is washed in 70% ethanol, dried and resuspended in 25 μ l TE. Sequencing was performed using a SequenaseTM sequencing kit following the protocol supplied by the manufacturer (U.S. Biochemical, Cleveland, OH).

EXAMPLE II

15 Isolation and Characterization of Peptide Ligands Generated From Oligonucleotides Having Random Codons at Two Predetermined Positions

This example shows the generation of a surface expression library from a population of oligonucleotides having randomized codons. The oligonucleotides are ten codons in length and are cloned into a single vector species for the generation of a M13 gene VIII-based surface expression library. The example also shows the selection of peptides for a ligand binding protein and characterization of their encoded nucleic acid sequences.

Oligonucleotide Synthesis

Oligonucleotides were synthesized as described in Example I. The synthesizer was programmed to synthesize the sequences shown in Table IX. These sequences correspond to the first random codon position synthesized and 3' flanking sequences of the oligonucleotide which hybridizes to the leader sequence in the vector. The

complementary sequences are used for insertional mutagenesis of the synthesized population of oligonucleotides.

Table IX

	<u>Column</u>	<u>Sequence (5' to 3')</u>
5	column 1	AA (A/C) GGTTCGGTCGGTACCGG
	column 2	AG (A/G) GGTTCGGTCGGTACCGG
	column 3	AT (A/G) GGTTCGGTCGGTACCGG
	column 4	AC (A/G) GGTTCGGTCGGTACCGG
10	column 5	CA (G/T) GGTTCGGTCGGTACCGG
	column 6	CT (G/C) GGTTCGGTCGGTACCGG
	column 7	AG (T/C) GGTTCGGTCGGTACCGG
	column 8	AT (T/C) GGTTCGGTCGGTACCGG
	column 9	CC (A/C) GGTTCGGTCGGTACCGG
15	column 10	T (A/T) TGGTTCGGTCGGTACCGG

The next eight random codon positions were synthesized as described for Table V in Example I. Following the ninth position synthesis, the reaction products were once more combined, mixed and redistributed into 10 new reaction columns. Synthesis of the last random codon position and 5' flanking sequences are shown in Table X.

Table X

	<u>Column</u>	<u>Sequence (5' to 3')</u>
25	column 1	AGGATCCGCCGAGCTCAA (A/C) A
	column 2	AGGATCCGCCGAGCTCAG (A/G) A
	column 3	AGGATCCGCCGAGCTCAT (A/G) A
	column 4	AGGATCCGCCGAGCTCAC (A/G) A
	column 5	AGGATCCGCCGAGCTCCA (G/T) A
	column 6	AGGATCCGCCGAGCTCCT (G/C) A
30	column 7	AGGATCCGCCGAGCTCAG (T/C) A
	column 8	AGGATCCGCCGAGCTCAT (T/C) A
	column 9	AGGATCCGCCGAGCTCCC (A/C) A
	column 10	AGGATCCGCCGAGCTCT (A/T) TA

The reaction products were mixed once more and the oligonucleotides cleaved and purified as recommended by the manufacturer. The purified population of oligonucleotides were used to generate a surface expression library as described below.

Vector Construction

The vector used for generating surface expression libraries from a single oligonucleotide population (i.e., without joining together of right and left half oligonucleotides) is described below. The vector is a M13-based expression vector which directs the synthesis of gene VIII-peptide fusion proteins (Figure 4). This vector exhibits all the functions that the combined right and left half vectors of Example I exhibit.

An M13-based vector was constructed for the cloning and surface expression of populations of random oligonucleotides (Figure 4, M13IX30), M13mp19 (Pharmacia) was the starting vector. This vector was modified to contain, in addition to the encoded wild type M13 gene VIII: (1) a pseudo-wild type gene, gene VIII sequence with an amber stop codon placed between it and the restriction sites for cloning oligonucleotides; (2) Stu I, Spe I and Xho I restriction sites in frame with the pseudo-wild type gVIII for cloning oligonucleotides; (3) sequences necessary for expression, such as a promoter, signal sequence and translation initiation signals; (4) various other mutations to remove redundant restriction sites and the amino terminal portion of Lac Z.

Construction of M13IX30 was performed in four steps. In the first step, a precursor vector containing the pseudo gene VIII and various other mutations was constructed, M13IX01F. The second step involved the construction of a small cloning site in a separate M13mp18 vector to yield

M13IX03. In the third step, expression sequences and cloning sites were constructed in M13IX03 to generate the intermediate vector M13IX04B. The fourth step involved the incorporation of the newly constructed sequences from the intermediate vector into M13IX01F to yield M13IX30. Incorporation of these sequences linked them with the pseudo gene VIII.

Construction of the precursor vector M13IX01F was similar to that of M13IX42 described in Example I except for the following features: (1) M13mp19 was used as the starting vector; (2) the Fok I site 5' to the unique Eco RI site was not incorporated and the overhang at the naturally occurring Fok I site at position 3547 was not changed to 5'-CTTC-3'; (3) the spacer sequence was not incorporated between the Eco RI and Sac I sites; and (4) the amber codon at position 4492 was not incorporated.

In the second step, M13mp18 was mutated to remove the 5' end of Lac Z up to the Lac i binding site and including the Lac Z ribosome binding site and start codon. Additionally, the polylinker was removed and a Mlu I site was introduced in the coding region of Lac Z. A single oligonucleotide was used for these mutagenesis and had the sequence "5'-AAACGACGGCCAGTGCCAAGTGACGCGTGTGAAATTGTTATCC-3'" (SEQ ID NO: 41). Restriction enzyme sites for Hind III and Eco RI were introduced downstream of the MluI site using the oligonucleotide "5'-GGCGAAAGGGAATTCTGCAAGGCGATTAAGCTTGGGTAACGCC-3'" (SEQ ID NO: 42). These modifications of M13mp18 yielded the vector M13IX03.

The expression sequences and cloning sites were introduced into M13IX03 by chemically synthesizing a series of oligonucleotides which encode both strands of the desired sequence. The oligonucleotides are presented in Table XI (SEQ ID NOS: 43 through 50).

TABLE XI
M13IX30 Oligonucleotide Series

<u>Top Strand</u> <u>Oligonucleotides</u>		<u>Sequence (5' to 3')</u>
5	084	GGCGTTACCCAAGCTTTGTACATGGAGAAAATAAAG
	027	TGAAACAAAGCACTATTGCACTGGCACTCTTACCGT TACCGT
	028	TACTGTTTACCCCTGTGACAAAAGCCGCCAGGTCC AGCTGC
10	029	TCGAGTCAGGCCTATTGTGCCCAGGGATTGTACTAG TGGATCCG
<u>Bottom</u> <u>Oligonucleotides</u>		<u>Sequence (5' to 3')</u>
	085	TGGCGAAAGGGAATTCGGATCCACTAGTACAATCCCTG
15	031	GGCACAAATAGGCCTGACTCGAGCAGCTGGACCAGGGCG GCTT
	032	TTGTCACAGGGGTAAACAGTAACGGTAACGGTAAGTGT GCCA
20	033	GTGCAATAGTGCTTTGTTTCACTTTATTTTCTCCATGT ACAA

The above oligonucleotides except for the terminal oligonucleotides 084 (SEQ ID NO: 43) and 085 (SEQ ID NO: 47) of Table XI were mixed, phosphorylated, annealed and ligated to form a double stranded insert as described in Example I. However, instead of cloning directly into the intermediate vector the insert was first amplified by PCR using the terminal oligonucleotides 084 (SEQ ID NO: 43) and 085 (SEQ ID NO: 47) as primers. The terminal oligonucleotide 084 (SEQ ID NO: 43) contains a Hind III site 10 nucleotides internal to its 5' end. Oligonucleotide 085 (SEQ ID NO: 47) has an Eco RI site at its 5' end. Following amplification, the products were restricted with Hind III and Eco RI and ligated as described in Example I into the polylinker of M13mpl8

digested with the same two enzymes. The resultant double stranded insert contained a ribosome binding site, a translation initiation codon followed by a leader sequence and three restriction enzyme sites for cloning random oligonucleotides (Xho I, Stu I, Spe I). The vector was named M13IX04.

During cloning of the double-stranded insert, it was found that one of the GCC codons in oligonucleotides 028 and its complement in 031 was deleted. Since this deletion did not affect function, the final construct is missing one of the two GCC codons. Additionally, oligonucleotide 032 contained a GTG codon where a GAG codon was needed. Mutagenesis was performed using the oligonucleotide 5'-TAACGGTAAGAGTGCCAGTGC-3' (SEQ ID NO: 51) to convert the codon to the desired sequence. The resultant intermediate vector was named M13IX04B.

The fourth step in constructing M13IX30 involved inserting the expression and cloning sequences from M13IX04B upstream of the pseudo-wild type gVIII in M13IX01F. This was accomplished by digesting M13IX04B with Dra III and Bam HI and gel isolating the 700 base pair insert containing the sequences of interest. M13IX01F was likewise digested with Dra III and Bam HI. The insert was combined with the double digested vector at a molar ratio of 3:1 and ligated as described in Example I. It should be noted that all modifications in the vectors described herein were confirmed by sequence analysis. The sequence of the final construct, M13IX30, is shown in Figure 7 (SEQ ID NO: 3). Figure 4 also shows M13IX30 where each of the elements necessary for surface expression of randomized oligonucleotides is marked.

Library Construction, Screening and Characterization of
Encoded Oligonucleotides

Construction of an M13IX30 surface expression library is accomplished identically to that described in Example I for sublibrary construction except the oligonucleotides described above are inserted into M13IX30 by mutagenesis instead of by ligation. The library is constructed and propagated on MK30-3 (BMB) and phage stocks are prepared for infection of XLI cells and screening. The surface expression library is screened and encoding oligonucleotides characterized as described in Example I.

EXAMPLE III

Isolation and Characterization of Peptide Ligands
Generated from Right and Left Half
Degenerate Oligonucleotides

15

This example shows the construction and expression of a surface expression library of degenerate oligonucleotides. The encoded peptides of this example derive from the mixing and joining together of two separate oligonucleotide populations. Also demonstrated is the isolation and characterization of peptide ligands and their corresponding nucleotide sequence for specific binding proteins.

Synthesis of Oligonucleotide Populations

25

A population of left half degenerate oligonucleotides and a population of right half degenerate oligonucleotides was synthesized using standard automated procedures as described in Example I.

The degenerate codon sequences for each population of oligonucleotides were generated by sequentially

30

synthesizing the triplet NNG/T where N is an equal mixture of all four nucleotides. The antisense sequence for each population of oligonucleotides was synthesized and each population contained 5' and 3' flanking sequences complementary to the vector sequence. The complementary termini was used to incorporate each population of oligonucleotides into their respective vectors by standard mutagenesis procedures. Such procedures have been described previously in Example I and in the Detailed Description. Synthesis of the antisense sequence of each population was necessary since the single-stranded form of the vectors are obtained only as the sense strand.

The left half oligonucleotide population was synthesized having the following sequence: 5'-AGCTCCCGGATGCCTCAGAAGATG(A/CNN)₉GGCTTTTGCCACAGGGG-3' (SEQ ID NO: 52). The right half oligonucleotide population was synthesized having the following sequence: 5'-CAGCCTCGGATCCGCC(A/CNN)₁₀ATG(A/C)GAAT-3' (SEQ ID NO. 53). These two oligonucleotide populations when incorporated into their respective vectors and joined together encode a 20 codon oligonucleotide having 19 degenerate positions and an internal predetermined codon sequence.

Vector Construction

Modified forms of the previously described vectors were used for the construction of right and left half sublibraries. The construction of left half sublibraries was performed in an M13-based vector termed M13ED03. This vector is a modified form of the previously described M13IX30 vector and contains all the essential features of both M13IX30 and M13IX22. M13ED03 contains, in addition to a wild type and a pseudo-wild type gene VIII, sequences necessary for expression and two Fok I sites for joining with a right half oligonucleotide

sublibrary. Therefore, this vector combines the advantages of both previous vectors in that it can be used for the generation and expression of surface expression libraries from a single oligonucleotide population or it can be joined with a sublibrary to bring together right and left half oligonucleotide populations into a surface expression library.

M13ED03 was constructed in two steps from M13IX30. The first step involved the modification of M13IX30 to remove a redundant sequence and to incorporate a sequence encoding the eight amino-terminal residues of human β -endorphin. The leader sequence was also mutated to increase secretion of the product.

During construction of M13IX04 (an intermediate vector to M13IX30 which is described in Example II), a six nucleotide sequence was duplicated in oligonucleotide 027 (SEQ ID NO: 44) and its complement 032 (SEQ ID NO: 49). This sequence, 5'-TTACCG-3', was deleted by mutagenesis in the construction of M13ED01. The oligonucleotide used for the mutagenesis was 5'-GGTAAACAGTAACGGTAAGAGTGCCAG-3' (SEQ ID NO: 54). The mutation in the leader sequence was generated using the oligonucleotide 5'-GGGCTTTTGCCACAGGGGT-3' (SEQ ID NO: 55). This mutagenesis resulted in the A residue at position 6353 of M13IX30 being changed to a G residue. The resultant vector was designated M13IX32.

To generate M13ED01, the nucleotide sequence encoding β -endorphin (8 amino acid residues of β -endorphin plus 3 extra amino acid residues) was incorporated after the leader sequence by mutagenesis. The oligonucleotide used had the following sequence: 5'-AGGGTCATCGCCTTCAGCTCCGGATCCCTCAGAAGTCATAAACCCCCCATAGGC TTTTGCCAC-3' (SEQ ID NO: 56). This mutagenesis also removed some of the downstream sequences through the Spe

I site.

The second step in the construction of M13ED03 involved vector changes which put the β -endorphin sequence in frame with the downstream pseudo-gene VIII sequence and incorporated a Fok I site for joining with a sublibrary of right half oligonucleotides. This vector was designed to incorporate oligonucleotide populations by mutagenesis using sequences complementary to those flanking or overlapping with the encoded β -endorphin sequence. The absence of β -endorphin expression after mutagenesis can therefore be used to measure the mutagenesis frequency. In addition to the above vector changes, M13ED03 was also modified to contain an amber codon at position 3262 for biological selection during joining of right and left half sublibraries.

The mutations were incorporated using standard mutagenesis procedures as described in Example I. The frame shift changes and Fok I site were generated using the oligonucleotide 5'-TCGCCTTCAGCTCCCGGATGCCTCAGAAGCATGAACCCCCCATAGGC-3' (SEQ ID NO: 57). The amber codon was generated using the oligonucleotide 5'-CAATTTTATCCTAAATCTTACCAAC-3' (SEQ ID NO: 58). The full sequence of the resultant vector, M13ED03, is provided in Figure 8 (SEQ ID NO: 4).

The construction of right half oligonucleotide sublibraries was performed in a modified form of the M13IX42 vector. The new vector, M13IX421, is identical to M13IX42 except that the amber codon between the Eco RI-SacI cloning site and the pseudo-gene VIII sequence was removed. This change ensures that all expression off of the Lac Z promoter produces a peptide-gene VIII fusion protein. Removal of the amber codon was performed by mutagenesis using the following oligonucleotide: 5'-GCCTTCAGCCTCGGATCCGCC-3' (SEQ ID NO: 59). The full

sequence of M13IX421 is shown in Figure 9 (SEQ ID NO: 5).

Library Construction, Screening and Characterization of
Encoded Oligonucleotides

A sublibrary was constructed for each of the
5 previously described degenerate populations of
oligonucleotides. The left half population of
oligonucleotides was incorporated into M13ED03 to
generate the sublibrary M13ED03.L and the right half
population of oligonucleotides was incorporated into
10 M13IX421 to generate the sublibrary M13IX421.R. Each of
the oligonucleotide populations were incorporated into
their respective vectors using site-directed mutagenesis
as described in Example I. Briefly, the nucleotide
sequences flanking the degenerate codon sequences were
15 complementary to the vector at the site of incorporation.
The populations of nucleotides were hybridized to single-
stranded M13ED03 or M13IX421 vectors and extended with T4
DNA polymerase to generate a double-stranded circular
vector. Mutant templates were obtained by uridine
20 selection in vivo, as described by Kunkel et al., supra.
Each of the vector populations were electroporated into
host cells and propagated as described in Example I.

The random joining of right and left half
sublibraries into a single surface expression library was
25 accomplished as described in Example I except that prior
to digesting each vector population with Fok I they were
first digested with an enzyme that cuts in the unwanted
portion of each vector. Briefly, M13ED03.L was digested
with Bgl II (cuts at 7094) and M13IX421.R was digested
30 with Hind III (cuts at 3919). Each of the digested
populations were further treated with alkaline
phosphatase to ensure that the ends would not religate
and then digested with an excess of Fok I. Ligations,
electroporation and propagation of the resultant library

was performed as described in Example I.

The surface expression library was screened for ligand binding proteins using a modified panning procedure. Briefly, 1 ml of the library, about 10^{12} phage particles, was added to 1-5 μ g of the ligand binding protein. The ligand binding protein was either an antibody or receptor globulin (Rg) molecule, Aruffo et al., Cell 61:1303-1313 (1990), which is incorporated herein by reference. Phage were incubated shaking with affinity ligand at room temperature for 1 to 3 hours followed by the addition of 200 μ l of latex beads (Biosite, San Diego, CA) which were coated with goat-antimouse IgG. This mixture was incubated shaking for an additional 1-2 hours at room temperature. Beads were pelleted for 2 minutes by centrifugation in a microfuge and washed with TBS which can contain 0.1% Tween 20. Three additional washes were performed where the last wash did not contain any Tween 20. The bound phage were then eluted with 200 μ l 0.1 M Glycine-HCl, pH 2.2 for 15 minutes and the beads were spun down by centrifugation. The supernatant-containing phage (eluate) was removed and phage exhibiting binding to the ligand binding protein were further enriched by one-to-two more cycles of panning. Typical yields after the first eluate were about 1×10^6 - 5×10^6 pfu. The second and third eluate generally yielded about 5×10^6 - 2×10^7 pfu and 5×10^7 - 1×10^{10} pfu, respectively.

The second or third eluate was plated at a suitable density for plaque identification screening and sequencing of positive clones (i.e., plated at confluency for rare clones and 200-500 plaques/plate if pure plaques were needed). Briefly, plaques grown for about 6 hours at 37°C and were overlaid with nitrocellulose filters that had been soaked in 2 mM IPTG and then briefly dried. The filters remained on the plaques overnight at room

temperature, removed and placed in blocking solution for 1-2 hours. Following blocking, the filters were incubated in 1 μ g/ml ligand binding protein in blocking solution for 1-2 hours at room temperature. Goat
5 antimouse Ig-coupled alkaline phosphatase (Fisher) was added at a 1:1000 dilution and the filters were rapidly washed with 10 mls of TBS or block solution over a glass vacuum filter. Positive plaques were identified after alkaline phosphatase development for detection.

10 Screening of the degenerate oligonucleotide library with several different ligand binding proteins resulted in the identification of peptide sequences which bound to each of the ligands. For example, screening with an
15 antibody to β -endorphin resulted in the detection of about 30-40 different clones which essentially all had the core amino acid sequence known to interact with the antibody. The sequences flanking the core sequences were different showing that they were independently derived and not duplicates of the same clone. Screening with an
20 antibody known as 57 gave similar results (i.e., a core consensus sequence was identified but the flanking sequences among the clones were different).

EXAMPLE IV

Generation of a Left Half Random Oligonucleotide Library

25 This example shows the synthesis and construction of a left half random oligonucleotide library.

A population of random oligonucleotides nine codons in length was synthesized as described in Example I except that different sequences at their 5' and 3' ends
30 were synthesized so that they could be easily inserted into the vector by mutagenesis. Also, the mixing and dividing steps for generating random distributions of

reaction products was performed by the alternative method of dispensing equal volumes of bead suspensions. The liquid chosen that was dense enough for the beads to remain dispersed was 100% acetonitrile.

- 5 Briefly, each column was prepared for the first coupling reaction by suspending 22 mg (1 μ mole) of 48 μ mol/g capacity beads (Genta, San Diego, CA) in 0.5 mls of 100% acetonitrile. These beads are smaller than those described in Example I and are derivatized with a guanine
10 nucleotide. They also do not have a controlled pore size. The bead suspension was then transferred to an empty reaction column. Suspensions were kept relatively dispersed by gently pipetting the suspension during transfer. Columns were plugged and monomer coupling
15 reactions were performed as shown in Table XII.

Table XII

<u>Column</u>	<u>Sequence</u> <u>(5' to 3')</u>
column 1L	AA(A/C)GGCTTTTGCCACAGG
20 column 2L	AG(A/G)GGCTTTTGCCACAGG
column 3L	AT(A/G)GGCTTTTGCCACAGG
column 4L	AC(A/G)GGCTTTTGCCACAGG
column 5L	CA(G/T)GGCTTTTGCCACAGG
column 6L	CT(G/C)GGCTTTTGCCACAGG
25 column 7L	AG(T/C)GGCTTTTGCCACAGG
column 8L	AT(T/C)GGCTTTTGCCACAGG
column 9L	CC(A/C)GGCTTTTGCCACAGG
column 10L	T(A/T)GGCTTTTGCCACAGG

- 30 After coupling of the last monomer, the columns were unplugged as described previously and their contents were poured into a 1.5 ml microfuge tube. The columns were rinsed with 100% acetonitrile to recover any remaining beads. The volume used for rinsing was determined so

that the final volume of total bead suspension was about 100 μ l for each new reaction column that the beads would be aliquoted into. The mixture was vortexed gently to produce a uniformly dispersed suspension and then divided, with constant pipetting of the mixture, into equal volumes. Each mixture of beads was then transferred to an empty reaction column. The empty tubes were washed with a small volume of 100% acetonitrile and also transferred to their respective columns. Random codon positions 2 through 9 were then synthesized as described in Example I where the mixing and dividing steps were performed using a suspension in 100% acetonitrile. The coupling reactions for codon positions 2 through 9 are shown in Table XIII.

15

Table XIII

	<u>Column</u>	<u>Sequence (5' to 3')</u>
	column 1L	AA(A/C)A
	column 2L	AG(A/G)A
20	column 3L	AT(A/G)A
	column 4L	AC(A/G)A
	column 5L	CA(G/T)A
	column 6L	CT(G/C)A
	column 7L	AG(T/C)A
25	column 8L	AT(T/C)A
	column 9L	CC(A/C)A
	column 10L	T(A/T)TA

After coupling of the last monomer for the ninth codon position, the reaction products were mixed and a portion was transferred to an empty reaction column. Columns were plugged and the following monomer coupling reactions were performed: 5'-CGGATGCCTCAGAAGCCCCXA-3' (SEQ ID NO: 60). The resulting population of random oligonucleotides was purified and incorporated by

30

mutagenesis into the left half vector M13ED04.

M13ED04 is a modified version of the M13ED03 vector described in Example III and therefore contains all the features of that vector. The difference between M13ED03 and M13ED04 is that M13ED04 does not contain the five amino acid sequence (Tyr Gly Gly Phe Met) recognized by anti- β -endorphin antibody. This sequence was deleted by mutagenesis using the oligonucleotide 5'-CGGATGCCTCAGAAGGGCTTTGCCACAGG (SEQ ID NO: 61). The entire nucleotide sequence of this vector is shown in Figure 10 (SEQ ID NO: 6).

Although the invention has been described with reference to the presently preferred embodiment, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Huse, William D.
- (ii) TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF
RANDOMIZED PEPTIDES
- (iii) NUMBER OF SEQUENCES: 61
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
 - (B) STREET: 444 South Flower Street, Suite 2000
 - (C) CITY: Los Angeles
 - (D) STATE: California
 - (E) COUNTRY: United States
 - (F) ZIP: 90071
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Campbell, Cathryn A
 - (B) REGISTRATION NUMBER: 31,815
 - (C) REFERENCE/DOCKET NUMBER: P31 9072
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (619) 535-9001
 - (B) TELEFAX: (619) 535-8949

(2) INFORMATION FOR SEQ ID NO:1:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: circular
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTGGCGCCCC AAATGAAAAT	60
ATAGCTAAAC AGGTTATTGA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCGCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG	360
TCTTTCCGGC TTCCTCTTAA TCTTTTGTAT GCAATCCGCT TTGCTTCTGA CTATAAGT	420

CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACT GTTTAAAGCA 480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT 540
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT 600
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT 660
AATTCCTTTT GCGGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG 720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT 780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA 840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT 900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG 960
AATATCCGGT TCTTGCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC 1020
TGACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC 1080
GTCTGCGCCT CGTCCGGCT AAGTAACATG GAGCAGGTG CGGATTTCA CACAATTTAT 1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAAT CGCTGGGGGT 1200
CAAAGATGAG TGTTTTAGTG TATTCTTTCG CCTCTTTCGT TTAGGTTGG TGCCTTCGTA 1260
GTGGCATTAC GTATTTTACC CGTTTAATGG AAACCTTCCTC ATGAAAAAGT CTTTAGTCTT 1320
CAAAGCCTCT GTAGCCGTG CTACCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA 1380
CGATCCCGCA AAAGCGGCCT TTAACCTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTGA 1440
TCCGTGGGCG ATGTTGTGTT TCATTGTTCG CGCAACTATC GGTATCAAGC TGTTTAAGAA 1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTTT GGAGCCTTTT 1560
TTTTTGGA GA TTTTCAACGT GAAAAAATTA TTATTGCAA TTCCTTIAGT TGTTCTTTT 1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTAGCAA AACCCCATAC AGAAAATTCA 1680
TTTACTAACC TCTGGAAAGA CGACAAAAC TTAGATCGTT ACGCTAACTA TGAGGGTTGT 1740
CTGTGGAATG CTACAGGCGT TGTTGTTGT ACTGGTGACG AAATCACTG TTACGGTACA 1800
TGGGTTCTTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT 1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT 1920
ATTCGGGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGCCTGG TACTGAGCAA 1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT 2040
CAGAATAATA GGTTCGAAA TAGGCAGGGG GCATTAACTG TTTATACGGG CACTGTACT 2100
CAAGGCACTG ACCCGGTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG 2160
TATGACGCTT ACTGGAACGG TAAATTGAGA CACTGCGCTT TCCATTCTGG CTTTAAATGAA 2220
GATCCATTG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCTGTCAAT 2280
GCTGGCGGCG GCTCTGGTGG TGGTCTGGT GCGGCTCTG AGGTTGGTGG CTCTGAGGGT 2340
GGCGGTTCTG AGGTTGGCGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC TGTTCCGGT 2400
GATTTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGG CTATGACCGA AAATGCCGAT 2460

GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACCTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT	2640
TTAATGAATA ATTTCCGTCA ATATTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT	2700
TTTGCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	2760
TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACA TACTGCGTAA TAAGGACTCT TAATCATGCC AGTTCTTTTG GGTATTCCGT	2880
TATTATTGCG TTTCCTCGGT TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTCGGTAAG ATAGCTATTG CTATTTGATT GTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTTGTC GGTTATCTCT CTGATATTAG CGCTCAATTA CCCTCTGACT	3060
TTGTTCAAGG TGTTCACTTA ATTCTCCCGT CTAATGCGCT TCCCTGTTTT TATGTTATTC	3120
TCTCTGTAAA GGCTGCTATT TTCATTTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG	3180
ATTGGGATAA ATAATATGGC TGTTTTATTT GTAAGTGGCA AATTAGGCTC TGGAAAGACG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT	3300
CTTGATTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTGCTTG CTATTGGGCG CGGTAATGAT	3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGCGGTAC TTGGTTTAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT CTTCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
CGTTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT	3660
TTTGTCGGTA CTTTATATT CTTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTTT ATTCTTATTT AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATTA	3900
AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGTA AAAAGTTTTC ACGCGTTCTT	3960
TGCTTTGCGA TTGGATTGCG ATCAGCATT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCATAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAAATTAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTTCC	4200
ATTAAAAAGG TAATTCAAAT GAAATTGTTA AATGTAATTA ATTTTGTTTT CTTGATGTTT	4260
GTTTCATCAT CTTCTTTTGC TCAGGTAATT GAAATGAATA ATTGCGCTCT GCGCGATTTT	4320
GTAAGTTGGT ATTCAAAGCA ATCAGGCGAA TCCGTTATTG TTTCTCCCGA TGTAAGAGGT	4380
ACTGTTACTG TATATTCATC TGACGTAAAA CCTGAAAATC TACGCAATTT CTTTATTTCT	4440
GTTTTACCTG CTAATAATTT TGATATGGTT GGTTCATTC CTTCCATTAT TTAGAAGTAT	4500

AATCCAAACA ATCAGGATTA TATTGATGAA TTGCCATCAT CTGATAATCA GGAATATGAT 4560
GATAATTCCG CTCCTTCTGG TGGTTTCTTT GTTCCGCAAA ATGATAATGT TACTCAAAC 4620
TTTAAAATTA ATAACGTTGG GCCAAAGGAT TTAATACCAG TTGTCGAATT GTTTGTAAAG 4680
TCTAATACTT CTAAATCCTC AAATGTATTA TCTATTGACG GCTCTAATCT ATTAGTTGTT 4740
AGTGCACCTA AAGATATTTT AGATAACCTT CCTCAATTCC TTTCTACTGT TGATTTGCCA 4800
ACTGACCAGA TATTGATTGA GGGTTTGATA TTTGAGGTTG AGCAAGGTGA TGCTTTAGAT 4860
TTTTCATTTG CTGCTGGCTC TCAGCGTGGC ACTGTTGCAG GCGGTGTTAA TACTGACCGC 4920
CTCACCTCTG TTTTATCTTC TGCTGGTGGT TCGTTGCGTA TTTTAAATGG CGATGTTTTA 4980
GGGCTATCAG TTCGCGCATT AAAGACTAAT AGCCATTCAA AAATATTGTC TGTGCCACGT 5040
ATTCTTACGC TTTGAGGTCA GAAGGGTTCT ATCTCTGTTG GCCAGAATGT CCCTTTTATT 5100
ACTGGTCGTG TGA CTGGTGA ATCTGCCAAT GTAAATAATC CATTTGAGAC GATTGAGCGT 5160
CAAAATGTAG GTATTTCAT GAGCGTTTTT CCTGTTGCAA TGGCTGGCGG TAATATTGTT 5220
CTGGATATTA CCAGCAAGGC CGATAGTTTG AGTCTTCTA CTCAGGCAAG TGATGTTATT 5280
ACTAATCAAA GAAGTATTGC TACAACGGTT AATTGCGGTG ATGGACAGAC TCTTTTACTC 5340
GGTGGCCTCA CTGATTATAA AAACACTTCT CAAGATTCTG GCGTACCGTT CCTGTCTAAA 5400
ATGCCCTTAA TCGGCCTCCT GTTTAGCTCC CGCTCTGATT CCAACGAGGA AAGCACGTTA 5460
TACGTGCTCG TCAAAGCAAC CATAGTAGGC GCCCTGTAGC GGCGCATTAA GCGCGGCGGG 5520
TGTTGGTGGT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGTCTCTTT 5580
CGCTTTCTTC CCTTCCTTTC TCGCCACGTT CGCCGGCTTT CCCCCTCAAG CTCTAAATCG 5640
GGGGCTCCCT TTAGGGTTCC GATTIAGTGC TTTACGGCAC CTCGACCCCA AAAAATTGA 5700
TTTGGGTGAT GGTTCACGTA GTGGGCCATC GCCCTGATAG ACGGTTTTTC GCCCTTTGAC 5760
GTGGAGTCC ACGTTCTTTA ATAGTGGACT CTGTTCCAA ACTGGAACAA CACTCAACCC 5820
TATCTCGGGC TATTCTTTTG ATTTATAAGG GATTTTGCCG ATTTCCGAAC CACCATCAAA 5880
CAGGATTTTC GCCTGCTGGG GCAAACCAGC GTGGACCGCT TGCTGCAACT CTCTCAGGGC 5940
CAGGCGGTGA AGGGCAATCA GCTGTTGCCC GTCTCGCTGG TGAAGAGAAA AACCACCCTG 6000
GCGCCCAATA CGGAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAAT GCAGCTGGCA 6060
CGACAGGTTT CCGGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAAATG TGAGTTAGCT 6120
CACTCATTAG GCACCCACAG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT 6180
TGTGAGCGGA TAACAATTTT ACACAGGAAA CAGCTATGAC CAGGATGTAC GAATTGCGAG 6240
GTAGGAGAGC TCGGCGGATC CTAGGCTGAA GGCGATGACC CTGCTAAGGC TGCATTCAAT 6300
AGTTTACAGG CAAGTGCTAC TGAGTACATT GGCTACGCTT GGGCTATGGT AGTAGTTATA 6360
GTGGGTGCTA CCATAGGGAT TAAATTATTC AAAAAGTTTA CGAGCAAGGC TTCTTAACCA 6420
GCTGGCGTAA TAGCGAAGAG GCGCGCACCG ATCGCCCTTC CCAACAGTTG CGCAGCCTGA 6480
ATGGCGAATG GCGCTTTGCC TGGTTTCCGG CACCAGAAGC GGTGCCGGAA AGCTGGCTGG 6540

AGTGGGATCT TCCTGAGGCC GATACGGTCG TCGTCCCCTC AAACGGCAG ATGCACGGTT	6600
ACGATGGGCC GATCTACACC AACGTAACCT ATCCCATTAC GGTCAATCCG CCGTTTGTTT	6660
CCACGGAGAA TCCGACGGGT TGTTACTCGC TCACATTTAA TGTTGATGAA AGCTGGCTAC	6720
AGGAAGGCCA GACGCGAATT ATTTTGATG GCGTTCCTAT TGGTIAAAAA ATGAGCTGAT	6780
TTAACAAAAA TTTAACGCGA ATTTTAACAA AATATTAACG TTTACAATTT AAATATTTGC	6840
TTATACAATC TTCCTGTTTT TGGGGCTTTT CTGATTATCA ACCGGGGTAC ATATGATTGA	6900
CATGCTAGTT TTACGATTAC CGTTCATCGA TTCTCTTGTT TGCTCCAGAC TCTCAGGCAA	6960
TGACCTGATA GCCTTTGTAG ATCTCTCAAA AATAGCTACC CTCTCCGGCA TTAATTTATC	7020
AGCTAGAACG GTTGAATATC ATATTGATGG TGATTGACT GTCTCCGGCC TTTCTCACCC	7080
TTTTGAATCT TTACCTACAC ATTACTCAGG CATTGCATTT AAAATATATG AGGGTTCTAA	7140
AAATTTTTAT CCTTGGCTTG AAATAAAGGC TTCTCCGGCA AAAGTATTAC AGGGTCATAA	7200
TGTTTTTGGT ACAACCGATT TAGCTTTATG CTCTGAGGCT TTATTGCTTA ATTTTGCTAA	7260
TTCTTTGCCT TGCCTGTATG ATTTATTGGA CGTT	7294

(2) INFORMATION FOR SEQ ID NO:2:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7320 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCCGCCCCC AAATGAAAAT	60
ATAGCTAAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TGGAAATGAAA CTTCAGACA CCGTACTTTA	180
GTTGCATATT TAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG	360
TCTTTGGGGC TTCCTCTTAA TCTTTTGTAT GCAATCCGCT TTGCTTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACT GTTTAAAGCA	480
TTGAGGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGCAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT	600
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT	660
AATTCCTTTT GCGGTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA	840

CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAGG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGCGCCT CGTTCGGGCT AAGTAACATG GAGCAGGTCC CGGATTTGCA CACAATTTAT	1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTTAGTG TATTCTTTCC CCTCTTTCTG TTTAGGTTGG TGCCTTCGTA	1260
GTGGCATTAC GTATTTTACC CGTTTAATGG AAACCTTCCTC ATGAAAAAGT CTTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTTG CTACCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA	1380
CGATCCCGCA AAAGCGGCCT TTAACCTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGCGTGCGCG ATGGTIGTTG TCATTGTCCG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTTT GGAGCCTTTT	1560
TTTTTGAGAG TTTTCAACGT GAAAAATTA TTATTGCGAA TTCCTTTAGT TGTTCTTTTC	1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTAGCAA AACCCCATAC AGAAAATTCA	1680
TTTACTAAGC TCTGGAAAGA CGACAAACT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGCGT TGTAGTTTGT ACTGGTGACC AAATCAGTG TTACGGTACA	1800
TGGGTTCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT	1920
ATTCCGGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGCCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTATGTTT	2040
CAGAATAATA GGTTCCGAAA TAGGCAGGGG GCATTAACTG TTTATACGGG CACTGTTACT	2100
CAAGGCACTG ACCCCGTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACGCTT ACTGGAACGG TAAATTGAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA	2220
GATCCATTCTG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACG TCCTGTCAAT	2280
GCTGGCGGCG GCTCTGGTGG TGTTCTGGT GCGGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GCGGTTCCG GTGGTGGCTC TGGTCCGGT	2400
GATTTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGG CTATGACCGA AAATGCCGAT	2460
GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACCTTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT	2640
TTAATGAATA ATTTCCGTCA ATATTIACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT	2700
TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	2760
TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATTCGGT	2880

TATTATTGCG TTTCTCGGT TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC 2940
TTAAAAAGGG CTTGCGTAAG ATAGCTATTG CTATTTCATT GTTCTTGCT CTTATTATTG 3000
GGCTTAACTC AATTCTTGTG GGTATCTCT CTGATATTAG CGCTCAATTA CCCTCTGACT 3060
TTGTTACAGG TGTTACAGTA ATTCTCCGT CTAATGGCT TCCCTGTTTT TATGTTATTC 3120
TCTCTGTAAA GGCTGCTATT TTCATTTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG 3180
ATTGGGATAA ATAATATGGC TGTTTATTTT GTAAGTGGCA AATTAGGCTC TGGAAAGACG 3240
CTCGTTAGCG TTGGTAAGAT TTAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT 3300
CTTGATTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAC GCCTCGCGTT 3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTGCTTG CTATTGGCG CGGTAATGAT 3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGCGGTAC TTGGTTAAT 3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTTCT ACATGCTCGT 3540
AAATTAGGAT GGGATATTAT CTTCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG 3600
CGTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT 3660
TTTGTCGGTA CTTTATATT CTTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT 3720
GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTAT 3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAACAGG CTTTTCTAG TAATTATGAT 3840
TCCGGTGTTT ATTCTTATT AACCCTTAT TTATCACAGG GTCGGTATT CAAACCATTA 3900
AATTTAGGTC AGAAGATGAA ATTAATAAA ATATATTGA AAAAGTTTT CCGGTTCTT 3960
TGTCTTGCGA TTGGATTGC ATCAGCATT ACATATAGTT ATATAACCCA ACCTAAGCCG 4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTGATA AATTCATAT TGAATCTTCT 4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGAAA ATTAATTAAT 4140
AGCGACGATT TACAGAAGCA AGGTATTCA CTCACATATA TTGATTATG TACTGTTTCC 4200
ATTAAAAAG GTAATTCAA TGAAATTGTT AAATGTAATT AATTTGTTT TCTTGATGTT 4260
TGTTTCATCA TCTCTTTT CTGAGTAAT TGAAATGAAT AATTCGCTC TCGCGATT 4320
TGTAAGTTGG TATTCAAAGC AATCAGGCG ATCCGTTATT GTTCTCCCG ATGTAAAAGG 4380
TACTGTTACT GTATATTCAT CTGACGTAA ACCTGAAAAT CTACGCAAT TCTTATTT 4440
TGTTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCCATAA TTCAGAAGTA 4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA 4560
TGATAATTCC GTCCTTCTG GTGGTTTCTT TGTTCCGCAA AATGATAATG TTAATCAAAC 4620
TTTTAAATT AATAACGTT GGGCAAAGGA TTTAATACGA GTTGTGCAAT TGTTGTAAA 4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT 4740
TAGTGCACCT AAAGATATT TAGATAACCT TCCTCAATTC CTTTCTACTG TTGATTGCC 4800
AACTGACCAG ATATTGATT AGGGTTTGAT ATTTGAGGT CAGCAAGGTG ATGCTTAGA 4860
TTTTTCATT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GCGGTGTTA ATACTGACCG 4920

CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTCGTTCCGT ATTTTAAATG GCGATGTTTT 4980
AGGGCTATCA GTTCGCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG 5040
TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTTAT 5100
TACTGGTCGT GTGACTGGTG AATCTGCCAA TGTAATAAT CCATTTCAGA CGATTGAGCG 5160
TCAAAATGTA GGTATTTCCA TGAGCGTTTT TCCTGTGCA ATGGCTGGCG GTAATATTGT 5220
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTCTTCT ACTCAGGCAA GTGATGTTAT 5280
TACTAATCAA AGAAGTATTG CTACAACGGT TAATTTGCGT GATGGACAGA CTCTTTTACT 5340
CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCCTACCGT TCCTGTCTAA 5400
AATCCCTTAA ATCGGCCTCC TGTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT 5460
ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCGCATT ACGCGGCGG 5520
GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCGGCTCCTT 5580
TCGCTTTCTT CCCTTCCTTT CTGCCCAGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC 5640
GGGGGCTCCC TTTAGGGTTC CGATTAGTG CTTTACGGCA CCTCGACCCC AAAAACTTG 5700
ATTTGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTGA 5760
CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTCCA AACTGGAACA AACTCAACC 5820
CTATCTCGG CTATTCTTTT GATTATAAG GGATTTTGGC GATTTCGAA CCACCATCAA 5880
ACAGGATTTT CGCCTGCTGG GGCAAACCAG CGTGGACCGC TTGCTGCAAC TCTCTCAGGG 5940
CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCCT 6000
GGGCCCCAAT ACGCAAACCG CCTCTCCCCG CGGTTGGCC GATTCAATTA TGCAGCTGGC 6060
ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGACTTAGC 6120
TCACTCATTA GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGAA 6180
TTGTGAGCGG ATAACAATT CACACGCCAA GGAGACAGTC ATAATGAAAT ACCTATTGCC 6240
TACGGCAGCC GCTGGATTGT TATTACTCGC TGCCGAACCA GCCATGGCCG AGCTCGTGAT 6300
GACCCAGACT CCAGAATTCC ATCCGGAATG AGTGTTAATT CTAGAACGGC TAAGCTTGGC 6360
ACTGGCCGTC GTTTTACAAC GTCGTGACTG GGAAAACCTT GCGTTACCC AACTTAATCG 6420
CCTTGACGA CACCCCCCTT TCGCCAGCTG GCGTAATAGC GAAGAGGCCG GCACCGATCG 6480
CCCTTCCCAA CAGTTGCGCA GCCTGAATGG CGAATGGCGC TTTGCCGTGT TTCGGCACC 6540
AGAAGCGGTG CCGGAAAGCT GGCTGGAGTG CGATCTTCTT GAGGCCGATA CGGTCGTCTG 6600
CCCCCAAAC TGGCAGATGC ACGGTTACGA TGCGCCCATC TACACCAACG TAACCTATCC 6660
CATTACGGTC AATCCGCCGT TTGTTCCAC GGAGAATCCG ACGGGTGTG ACTCGCTCAC 6720
ATTTAATGTT GATGAAAGCT GGCTACAGGA AGCCAGACG CGAATTATTT TTGATGGCGT 6780
TCCTATTGGT TAAAAATGA GCTGATTAA CAAAAATTA ACGCGAATTT TAACAAAATA 6840
TTAACGTTTA CAATTTAAAT ATTTGCTTAT ACAATCTTCC TGTTTTGGG GCTTTTCTGA 6900
TTATCAACCG GGTACATAT GATTGACATG CTAGTTTAC GATTACCGTT CATCGATTCT 6960

CTTGTTTGCT CCAGACTCTC AGGCAATGAC CTGATAGCCT TTGTAGATCT CTCAAAAATA 7020
 GCTACCCTCT CCGGCATTAA TTTATCAGCT AGAACGGTTC AATATCATAT TGATGGTGAT 7080
 TTGACTGTCT CCGGCCCTTC TCACCCTTTT GAATCTTTAC CTACACATTG CTCAGGCATT 7140
 GCATTTAAAA TATATGAGGG TTCTAAAAAT TTTTATCCTT GCGTTGAAAT AAAGGCTTCT 7200
 CCCGCAAAAG TATTACAGGG TCATAATGTT TTTGGTACAA CCGATTTAGC TTTATGCTCT 7260
 GAGGCTTTAT TGCTTAATTT TGCTAATTCT TTGCCTTGCC TGTATGATT ATTGGACGTT 7320

(2) INFORMATION FOR SEQ ID NO:3:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7445 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGCGCCCC AAATGAAAAT 60
 ATAGCTAAAC AGTTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT 120
 CGTTCGCAGA ATTGGGAATC AACTGTTACA TGGAAATGAAA CTTCAGACA CCGTACTTTA 180
 GTTGCAATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA 240
 TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG 300
 TTGGAGTTTG CTTCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAACGCG ATATTGAAG 360
 TCTTCGGGC TTCCTCTTAA TCTTTTGTAT GCAATCCGCT TTGCTTCTGA CTATAATAGT 420
 CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACT GTTTAAAGCA 480
 TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT 540
 AAACATTTTA CTATTACCCC CTCTGGCAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT 600
 GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT 660
 AATTCCTTTT GCGGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG 720
 ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT 780
 TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA 840
 CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT 900
 CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG 960
 AATATCCGGT TCTTGTCAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC 1020
 TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CCGTTCCCTT ATGATTGACC 1080
 GTCTGCGCCT CGTTCGGGCT AAGTAACATG GAGCAGGTCC CGGATTTCTA CACAATTTAT 1140
 CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAAT CGCTGGGGGT 1200
 CAAAGATGAG TGTTTTAGTG TATTCTTTCG CCTCTTTCGT TTTAGGTGG TGCCTTCGTA 1260

GTGGCATTAC	GTATTTTACC	CGTTTAATGG	AACTTCCTC	ATGAAAAAGT	CTTTAGTCCT	1320
CAAAGCCTCT	GTAGCCGTG	CTACCCCTCGT	TCCGATGCTC	TCTTTCGCTG	CTGAGGGTGA	1380
CGATCCCGCA	AAAGCGGCCT	TAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440
TGGTGGGCG	ATGGTTGTTG	TCATTGTGGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA	1500
ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT	1560
TTTTTGAGA	TTTTCAACGT	GAAAAAATTA	TTATTGCGAA	TTCCTTTAGT	TGTTCCTTTC	1620
TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA	1680
TTTACTAACG	TCTGGAAAGA	CGACAAAAC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
CTGTGGAATG	CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA	1800
TGGGTTCCCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860
TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920
ATTCGGGGCT	ATACTTATAT	CAACCCTCTC	GACGGCACTT	ATCCGCCTGG	TACTGAGCAA	1980
AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	2040
CAGAATAATA	GGTTCCGAAA	TAGGCAGGGG	GCATTAAC	TTTATACGGG	CACTGTTACT	2100
CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG	2160
TATGAGGCTT	ACTGGAACGG	TAAATTCAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA	2220
GATCCATTGG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT	2280
GCTGGCGGCG	GCTCTGGTGG	TGGTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT	2400
GATTTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATCACCGA	AAATGCCGAT	2460
GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AACTTGATT	CTGTCCGTAC	TGATTACGGT	2520
GCTGCTATCG	ATGTTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	2580
GCTGATTTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCACCT	2640
TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT	2700
TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTTTCTATTG	ATTGTGACAA	AATAAACTTA	2760
TTCCGTGGTG	TCTTTGCGTT	TCTTTTATAT	GTGCCACCT	TTATGTATGT	ATTTTCTACG	2820
TTTGCTAACA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTTG	GGTATTCGGT	2880
TATTATTGCG	TTTCCCTGGT	TTCCTTCTGG	TAACCTTGTT	CGGCTATCTG	CTTACTTTTC	2940
TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCAAT	GTTTCTTGCT	CTTATTATTG	3000
GGCTTAACTC	AATTCTTG	GGTATCTCT	CTGATATTAG	CGCTCAATTA	CCCTCTGACT	3060
TTGTTCAAGG	TGTTCAAGTTA	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTTT	TATGTTATTC	3120
TCTCTGTAAA	GGCTGCTATT	TTGATTTTTG	ACGTAAACA	AAAAATCGTT	TCTTATTTGG	3180
ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAAGTGGCA	AATTAGGCTC	TGGAAAGACG	3240
CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT	3300

CTTGATTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTGCTTG CTATTGGGCG CGGTAATGAT	3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTCCGGTAC TTGGTTTAAT	3480
ACCCGTTCTT GGAATGATAA GCAAAGACAG CCGATTATTG ATTGGTTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT TTTTCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
CGTTCTGCAT TAGCTGAACA TGTGTTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT	3660
TTTGTGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTTCTAG TAATTATGAT	3840
TCCGGTGTIT ATTCTTATTT AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATT	3900
AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTTGA AAAAGTTTTT ACGCCGTCTT	3960
TGTCTTGCGA TTGGATTTCG ATCAGCATT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCATAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT	4140
AGCGACGATT TACAGAAGCA AGGTATTCA CTCACATATA TTGATTATG TACTGTTTCC	4200
ATTAAAAAAG GTAATTCAA TGAAATTGTT AAATGTAATT AATTTTGTIT TCTTGATGTT	4260
TGTTTCATCA TCTTCTTTTG CTCAGGTAAT TGAAATGAAT AATTCGCCCTC TCGCGGATTT	4320
TGTAACCTGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTCTCCCG ATGTAAAAGG	4380
TACTGTTACT GTATATTCAT CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTTATTTT	4440
TGTTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTCCATAA TTCAGAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCTTCTG GTGGTTTCTT TGTTCCGCAA AATGATAATG TTAATCAAAC	4620
TTTTAAATTT AATAACGTTT GGGCAAAGGA TTTAATACGA GTTGTGCAAT TGTTTGTAAT	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT	4740
TAGTGACCT AAAGATATTT TAGATAACCT TCCTCAATTC CTTTCTACTG TTGATTGCCC	4800
AACTGACCAG ATATTGATTG AGGTTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA	4860
TTTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTIA ATACTGACCG	4920
CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTCGTTGGT ATTTTAAATG GCGATGTTTT	4980
AGGGCTATCA GTTCGCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTTAT	5100
TACTGGTCGT GTGACTGGTG AATCTGCCAA TGTAATAAAT CCATTTTACA CGATTGAGCG	5160
TCAAAATGTA GGTATTTCCA TGAGCGTTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT	5220
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTCTTCT ACTCAGGCAA GTGATGTTAT	5280
TACTAATCAA AGAAGTATTG CTACAACGGT TAATTTGCGT GATGGACAGA CTCTTTTACT	5340

CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GCGGTACCGT TCCTGTCTAA	5400
AATCCCTTTA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT	5460
ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCGCATT ACGCGGCGG	5520
GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT	5580
TGCTTTCTT CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCGTCAA GCTCTAAATC	5640
GGGGGCTCCC TTTAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCGC AAAAACTTG	5700
ATTTGGGTGA TGTTTCAGGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA	5760
CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTTCCA AACTGGAACA AACTCAACC	5820
CTATCTGGG CTATTCTTTT GATTATAAG GGATTTTGCC GATTTCGGAA CCACCATCAA	5880
ACAGGATTTT CGCCTGCTGG GGCAAACCAG CGTGACCGC TTGCTGCAAC TCTCTCAGG	5940
CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCT	6000
GGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTA TGCAGCTGGC	6060
ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC	6120
TCACTCATT A GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA	6180
TTGTGAGCGG ATAACAATTT CACACGCGTC ACTTGGCACT GGCCGTCGTT TTACAACGTC	6240
GTGACTGGGA AAACCCTGGC GTTACCCAAG CTTGTACAT GGAGAAAATA AAGTGAACA	6300
AAGCACTATT GCACTGGCAC TCTTACCGTT ACCGTTACTG TTTACCCCTG TGACAAAAGC	6360
CGCCCAGGTC CAGCTGCTCG AGTCAGGCCT ATTGTGCCCA GGGGATTGTA CTAGTGGATC	6420
CTAGGCTGAA GGCGATGACC CTGCTAAGGC TGCATTGAAT AGTTTACAGG CAAGTGCTAC	6480
TGAGTACATT GGCTACGCTT GGGCTATGGT AGTAGTTATA GTTGGTGCTA CCATAGGGAT	6540
TAAATTATTC AAAAAGTTTA CGAGCAAGGC TTCTTAAGCA ATAGCGAAGA GGCCCGCACC	6600
GATCGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT GGCGCTTGC CTGGTTCCG	6660
GCACCAGAAG CGGTGCCGGA AAGCTGGCTG GAGTGGGATC TTCCTGAGGC CGATACGGTC	6720
GTGCTCCCCT CAAACTGGCA GATGCACGGT TACGATGCGC CCATCTACAC CAACGTAACC	6780
TATCCCATTA CGGTCAATCC GCCGTTTGTT CCCACGGAGA ATCCGACGGG TTGTTACTCG	6840
CTCAGATT A TGTTGATGA AAGCTGGCTA CAGGAAGGCC AGACGCGAAT TATTTTGTAT	6900
GGCGTTCCTA TTGGTTAAAA AATGAGCTGA TTAAACAAAA ATTAAACGCG AATTTTAAAC	6960
AAATATTAACT GTTTACAATT TAAATATTG CTTATACAAT CTTCTGTTT TTGGGGCTTT	7020
TCTGATTATC AACCGGGGTA CATATGATTG ACATGCTAGT TTTACGATTA CCGTTCATCG	7080
ATTCTCTTGT TTGCTCCAGA CTCTCAGGCA ATGACCTGAT AGCCTTTGTA GATCTCTCAA	7140
AAATAGCTAC CCTCTCCGGC ATTAATTTAT CAGCTAGAAC GGTGAATAT CATATTGATG	7200
GTGATTIGAC TGTCTCCGGC CTTTCTCACC CTTTGAATC TTACCTACA CATTACTCAG	7260
GCATTGCATT TAAAATATAT GAGGGTTCIA AAAATTTTIA TCCTTGCGTT GAAATAAAGG	7320
CTTCTCCGGC AAAAGTATTA CAGGGTCATA ATGTTTTTGG TACAACCGAT TTAGCTTTAT	7380

GCTCTGAGGC TTTATTGCTT AATTTTGCTA ATTCTTTGCC TTGCCTGTAT GATTIATTGG 7440
ACGTT 7445

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7409 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGCGCCCC AAATGAAAAT	60
ATAGCTAAAC AGGTTATTGA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG	360
TCTTCGGGC TTCCTCTTAA TCTTTTIGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTTT TGATTATGG TCATTCTCGT TTTCTGAACT GTTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGCAA ACTTCTTTG CAAAAGCCTC TCGCTATTTT	600
GTTTTTTATC GTCGCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT	660
AATTCCTTTT GCGGTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTAGTTC GTTTTATTAA CGTAGATTTT	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTCTTA AAATCGCATA AGGTAATTCA	840
CAATGATTAA AGTTGAAAT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CCGTTCCCTT ATGATTGACC	1080
GTCTGCGCCT CGTTCGGGCT AAGTAACATG GAGCAGGTG CCGATTTCGA CACAATTTAT	1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTCGCGC TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTTAGTG TATCTTTTGG CCTCTTTCGT TTTAGGTTGG TGCCTTCGTA	1260
GTGGCATTAC GTATTTTACC CGTTAATGG AAATCTCCTC ATGAAAAAGT CTTAGTCT	1320
CAAAGCCTCT GTAGCCGTTG CTACCCTCGT TCCGATGCTG TCTTCGCTG CTGAGGGTGA	1380
CGATCCCGCA AAAGCGGCCT TTAATCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGCGTGGGCG ATGGTTGTTG TCATTGTGCG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1500

ATTACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTTT GGAGCCTTTT	1560
TTTTTGAGA TTTTCAACGT GAAAAAATTA TTATTCGCAA TTCCTTTAGT TGTTCCTTTC	1620
TATTCTCACT CCGCTGAAAC TGTGAAAGT TGTTTAGCAA AACCCCATAC AGAAAAATCA	1680
TTTACTAACG TCTGAAAGA CGACAAAAC TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGCGT TGTAGTTTGT ACTGGTGACG AAACCTCAGT TTACGGTACA	1800
TGGGTTCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT	1920
ATTCCGGGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCGGCCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAATCC TTCTCTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT	2040
CAGAATAATA GGTTCGAAA TAGGCAGGGG GCATTAACTG TTTATACGGG CACTGTACT	2100
CAAGGCACTG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACGCTT ACTGGAACGG TAAATTCAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA	2220
GATCCATTCTG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCTGTCAAT	2280
GCTGGCGGCG GCTCTGGTGG TGTTCTGGT GCGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGCGGTTCTG AGGTGGCGG CTCTGAGGGA GCGGTTCCG GTGGTGCTC TGGTTCCGGT	2400
GATTTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2460
GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAAGTTGATT CTGTCCCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT	2640
TTAATGAATA ATTTCCGTCA ATATTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT	2700
TTTGTCTTTA GCGCTGCTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	2760
TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATTCGGT	2880
TATTATTGCG TTTCTCGGT TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTCATT GTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCCTTGTG GGTATCTCT CTGATATTAG CGCTCAATTA CCTCTGACT	3060
TTGTTACGGG TGTTCACTTA ATTCTCCCGT CTAATGCGCT TCCCTGTTTT TATGTTATTC	3120
TCTCTGTAAA GGCTGCTATT TTCATTTTGG ACGTTAAACA AAAATCGTT TCTTATTGG	3180
ATTGGGATAA ATAATATGGC TGTTTATTTT GTAAGTGGCA AATTAGGCTC TGGAAAGACG	3240
CTCGTTAGCG TTGTAAGAT TCAGGATAAA ATTGTAGCTG GTTGCAAAAT AGCAACTAAT	3300
CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTTGCTTG CTATTGGGCG CGGTAATGAT	3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTCCGGTAC TTGGTTTAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTTCT ACATGCTCGT	3540

AAATTAGGAT GGGATATTAT TTTTCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
CGTTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT	3660
TTTGTGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTTCTAG TAATTATGAT	3840
TCCGGTGTTT ATTCTTATTT AAGGCCTTAT TTATCACACG GTCGGTATTT CAAACCATTA	3900
AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTTGA AAAAGTTTTT ACGCCGTTCTT	3960
TGTCTTGCGA TTGGATTGTC ATCAGCATTT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCATAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTTCC	4200
ATTAAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTTTGTTT TCTTGATGTT	4260
TGTTTCATCA TCTTCTTTTG CTCAGGTAAT TGAAATGAAT AATTCGCCCTC TGCGCGATTT	4320
TGTAACITGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTCTCCCG ATGTAAAAGG	4380
TACTGTTACT GTATATTCAT CTGACGTAA ACCTGAAAAT CTACGCAATT TCTTTATTTT	4440
TGTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTCCATAA TTCAGAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCTTCTG GTGGTTTCTT TGTTCCGCAA AATGATAATG TTAATCAAAC	4620
TTTTAAAATT AATAACGTTT GGGCAAAGGA TTTAATACGA GTTGTGCAAT TGTTTGTAAG	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT	4740
TAGTGCACCT AAAGATATTT TAGATAACCT TCCTCAATTC CTTTCTACTG TTGATTTGCC	4800
AACTGACCAG ATATTGATTG AGGGTTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA	4860
TTTTTCAATT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GCGGCTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTCGTTCCGT ATTTTAAATG GCGATGTTTT	4980
AGGGCTATCA GTTCGCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTTAT	5100
TACTGGTCGT GTGACTGGTG AATCTGCCAA TGTAATAAAT CCATTTCAGA CGATTGAGCG	5160
TCAAAATGTA GGTATTTCCA TGAGCGTTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT	5220
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT	5280
TACTAATCAA AGAAGTATTG CTACAACGGT TAATTTGCGT GATGGACAGA CTCTTTTACT	5340
CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCTGTCTAA	5400
AATCCCTTTA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCAGGTT	5460
ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCGCATTA AGCGCGGCGG	5520
GTGTGGTGCT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCGTT	5580

TCGCTTTCTT CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC	5640
GGGGGCTCCC TTTAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAACTTG	5700
ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA	5760
CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTTCCTA AACTGGAACA AACTCAACC	5820
CTATCTCGGG CTATTCTTTT GATTTATAAG GGATTTTGCC GATTTCGGAA CCACCATCAA	5880
ACAGGATTTT CGCCTGCTGG GGCAAACCAG CGTGGACCGC TTGCTGCAAC TCTCTCAGGG	5940
CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCCT	6000
GGCGCCCAAT ACCGAAACCG CCTCTCCCGG CGCGTTGGCC GATTCATTAA TGCAGCTGGC	6060
ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC	6120
TCACTCATTG GGCAGCCAG GCTTTACACT TTAGCTTCC GGCTCGTATG TTGTGTGGAA	6180
TTGTGAGCGG ATAACAATT CACACGGTC ACTTGGCACT GGCCGTGGT TTACAACGTC	6240
GTGACTGGGA AAACCCTGGC GTTACCCAAG CTTTGTACAT GGAGAAAATA AAGTGAAACA	6300
AAGCACTATT GCACTGGCAC TCTTACCGTT ACTGTTTACC CCTGTGGCAA AAGCCTATGG	6360
GGGGTTTATG ACTTCTGAGG GATCCGGAGC TGAAGCGGAT GACCCTGCTA AGGCTGCATT	6420
CAATAGTTTA CAGGCAAGTG CTA CTGAGTA CATTGGCTAC GCTTGGGCTA TGGTAGTAGT	6480
TATACTTGGT GCTACCATAG GGATTAAATT ATTCAAAAAG TTTACGAGCA AGGCTTCTTA	6540
AGCAATAGCG AAGAGGCCCC CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC	6600
GAATGGCGCT TTGCCTGGTT TCCGGCACCA GAAGCGGTGC CGGAAAGCTG GCTGGAGTGC	6660
GATCTTCTGT AGGCCGATAC GGTCTCTCTC CCTCAAAC GGCAGATGCA CGGTTACGAT	6720
GGCCCCATCT ACACCAACGT AACCTATCCC ATTACGGTCA ATCCGCCGTT TGTTCCCACG	6780
GAGAATCCGA CGGGTTGTTA CTCGCTCACA TTTAATGTTG ATGAAAGCTG GCTACAGGAA	6840
GGCCAGACGC GAATTATTTT TGATGGCGTT CCTATTGGTT AAAAAATGAG CTGATTTAAC	6900
AAAAATTTAA CGCGAATTTT AACAAAATAT TAACGTTTAC AATTTAAATA TTGCTTATA	6960
CAATCTTCTT GTTTTTGGGG CTTTCTGAT TATCAACCGG GGTACATATG ATTGACATGC	7020
TAGTTTTIACG ATTACCGTTC ATCGATTCTC TTGTTTGCTC CAGACTCTCA GGCAATGACC	7080
TGATAGCCTT TGATAGATCTC TCAAAAATAG CTACCCTCTC CGGCATTAAT TTATCAGCTA	7140
GAACGGTTGA ATATCATATT GATGGTGATT TGACTGTCTC CGGCCTTTCT CACCCTTTTG	7200
AATCTTTACC TACACATTAC TCAGGCATTG CATTTAAAAAT ATATGAGGGT TCTAAAAATT	7260
TTTATCCTTG CGTTGAAATA AAGGCTCTC CGGCAAAAGT ATTACAGGCT CATAATGTTT	7320
TTGGTAGAAC CGATTIAGCT TTAGCTCTG AGGCTTIIATT GCTTAATTTT GCTAATTCTT	7380
TGCCTTGCCT GTATGATTTA TTGGACGTT	7409

(2) INFORMATION FOR SEQ ID NO:5:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7294 bas pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGCGCCCC AAATGAAAAT	60
ATAGCTAAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCGCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCGGGTCT GGTTGCTTTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG	360
TCTTCGGGGC TTCCTCTTAA TCTTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACT GTTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT	600
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT	660
AATTCCTTTT GCGGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAG ATTACTCTTG ATGAAGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CGGTTCGCTT ATGATTGACC	1080
GTCTGCGCCT CGTTCGGGCT AAGTAACATG GAGCAGGTCG CGGATTTCCA CACAATTTAT	1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGGCG TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTTAGTG TATTCTTTTCG CCTCTTTCGT TTTAGGTTGG TGCCTTCGTA	1260
GTGGCATTAC GTATTTTACC CGTTTAATGG AAATCTCCTC ATGAAAAAGT CTTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTTG CTACCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA	1380
CGATCCCGCA AAAGCGGCCT TTAATCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGCGTGGGCG ATGGTTGTTG TCATTGTGGG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGTCCTTTT GGAGCCTTTT	1560
TTTTTGAGAA TTTTCAACGT GAAAAAATTA TTATTCGCAA TTCCTTTAGT TGTTCCTTTC	1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTAGCAA AAGCCCATAC AGAAAATTCA	1680

TTTACTAACG TCTGGAAAGA CGACAAACT TTAGATCGTT ACGCTAACTA TGAGGGTTGT 1740
CTGTGGAATG CTACAGGCGT TGTAGTTTGT ACTGGTGACG AAACCTCAGTG TTACGGTACA 1800
TGGGTTCCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT 1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT 1920
ATTCCGGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGCCTGG TACTGAGCAA 1980
AACCCCGCTA ATCCTAATCC TTCTCTTGTAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT 2040
CAGAATAATA GGTTCGAAA TAGGCAGGGG GCATTAACCTG TTTATACGGG CACTGTACT 2100
CAAGGCACTG ACCCCGTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG 2160
TATGACGCTT ACTGGAACGG TAAATTCAGA GACTGCGCTT TCCATTCTGG CTTAATGAA 2220
GATCCATTGG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCTGTCAAT 2280
GCTGGCGGCG GCTCTGGTGG TGCTTCTGGT GGCGGCTCTG AGGGTGGTGG CTCTGAGGGT 2340
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC TGCTTCCGGT 2400
GATTTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT 2460
GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACCTGATT CTGTGCTAC TGATTACGGT 2520
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT 2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT 2640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTGA ATGTCCGCCT 2700
TTTGTCTTTA GCGCTGTAA ACCATATGAA TTTTCTATTG ATGTGACAA AATAAACTTA 2760
TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT TTATGTATGT ATTTTCTACG 2820
TTTGCTAACA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTCTTTTG GGTATTCCGT 2880
TATTATTCGG TTTCTCCGT TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC 2940
TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTTTATT GTTCTTGCT CTTATTATG 3000
GGCTTAACTC AATTCTTG TGTTATCTCT CTGATATTAG CGCTCAATTA CCCTCTGACT 3060
TTGTTACAGG TGTTCACTTA ATTCTCCCGT CTAATGCGCT TCCCTGTTTT TATGTTATTC 3120
TCTCTGIAAA GGCTGCTATT TTCATTTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG 3180
ATTGGGATAA ATAATATGGC TGTTTATTTT GTAAGTGGA AATTAGGCTC TGGAAAGACG 3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT 3300
CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAAC GCCTCGCGTT 3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTGTCTG CTATTGGGCG CGGTAATGAT 3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGCGGTAC TTGGTTAAT 3480
ACCCGTTCCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTTCT ACATGCTCCT 3540
AAATTAGGAT GGGATATTAT CTTCTTGTG CAGGACTTAT CTATTGTTGA TAAACAGGCG 3600
CGTTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCCTCGTC TGGACAGAAT TACTTTACCT 3660
TTTGTCCGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT 3720

GTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTTT ATTCTTATTT AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATTAA	3900
AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTTGA AAAAGTTTTT ACGCCCTTCT	3960
TGTCTTGCGA TTGGATTTCG ATCAGCATTT ACATAIAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCATAT TGA CTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGAAA ATTAATTAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATICA CTCACATATA TTGATTTATG TACTGTTTCC	4200
ATTAAAAAGG TAATTCAAAT GAAATGTGA AATGTAATTA ATTTGTTTT CTGATGTTT	4260
GTTTCATCAT CTCTTTTTGC TCAGGTAATT GAAATGAATA ATTCCCTCT GCGCGATTTT	4320
GTAAGTTGGT ATCAAAGCA ATCAGGCGAA TCCGTTATTG TTTCTCCGA TGTAAGGTT	4380
ACTGTTACTG TATATTCATC TGACGTTAAA CCTGAAATC TACGCAATTT CTTTATTTCT	4440
GTTTTACGTG CTAATAATTT TGATATGGTT GGTTCATTC CTTCATTAT TTAGAAGTAT	4500
AATCCAAACA ATCAGGATTA TATTGATGAA TTGCCATCAT CTGATAATCA GGAATATGAT	4560
GATAATTCCG CTCCTTCTGG TGGTTTCTTT GTTCCGAAA ATGATAATGT TACTCAAAT	4620
TTTAAATTA ATAACGTTG GGCAGGAT TTAATACGAG TTGTCGAATT GTTGTAAG	4680
TCTAATACTT CTAAATCTC AAATGTATTA TCTATTGACC GCTCTAATCT ATTAGTTGTT	4740
AGTGACCTA AAGATATTTT AGATAACCTT CCTCAATTC TTTCTACTGT TGATTTGCCA	4800
ACTGACCAGA TATTGATTGA GGGTTTGATA TTTGAGGTT AGCAAGGTGA TGCTTTAGAT	4860
TTTTCATTTG CTGCTGGCTC TCAGCGTGGC ACTGTGCAG GCGGTGTAA TACTGACCGC	4920
CTCACCTCTG TTTTATCTTC TGCTGGTGGT TCGTTCCGTA TTTTAAATGG CGATGTTTA	4980
GGGCTATCAG TTCGCGCATT AAAGACTAAT AGCCATTCAA AAATATTGTC TGTGCCACGT	5040
ATTCTTACGC TTTAGGTCA GAAGGGTTCT ATCTCTGTTG GCCAGAATGT CCGTTTTATT	5100
ACTGGTCGTG TGA CTGGTGA ATCTGCCAAT GTAAATAATC CATTTCAGAC GATTGACCGT	5160
CAAAATGTAG GTATTTCCAT GAGCGTTTTT CCTGTGCAA TGGCTGGCGG TAATATTGTT	5220
CTGGATATTA CCAGCAAGGC CGATAGTTTG AGTCTTCTA CTCAGGCAAG TGATGTTATT	5280
ACTAATCAAA GAAGTATTGC TACAACGGTT AATTGCGTG ATGGACAGAC TCTTTTACTC	5340
GGTGGCCTCA CTGATTATAA AAACACTTCT CAAGATTCTG GCGTACCGTT CCTGTCTAAA	5400
ATCCCTTTAA TCGGCCTCCT GTTTAGCTCC CGCTCTGATT CCAACGAGGA AAGCACGTTA	5460
TACGTGCTCG TCAAAGCAAC CATAGTACGC GCCCTGTAGC GGCGCATTA GCGCGGGGG	5520
TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT	5580
CGCTTTCTTC CTTTCCTTC TCGCCACGTT CGCCGGCTTT CCGGTCAAG CTCTAAATCG	5640
GCGGCTCCCT TTAGGGTTCC GATTTAGTGC TTTACGGCAC CTCGACCCCA AAAA ACTTGA	5700
TTTGGGTGAT GGTTCACGTA GTGGGCCATC GCCCTGATAG ACGGTTTTTC GCCCTTTGAC	5760

GTGGAGTCC	ACGTTCTTTA	ATAGTGGACT	CTTGTTCCAA	ACTGGAACAA	CACTCAACCC	5820
TATCTCGGGC	TATTCCTTTG	ATTATAAGG	GATTTTGCCG	ATTCGGAAC	CACCATCAAA	5880
CAGGATTTTC	GCCTGCTGGG	GCAAACCAGC	GTGGACCGCT	TGCTGCAACT	CTCTCAGGGC	5940
CAGGCGGTGA	AGGGCAATCA	GCTGTTGCCC	GTCTCGCTGG	TGAAAAGAAA	AACCACCCTG	6000
CGGCCCAATA	CGCAAACCGC	CTCTGCCCGC	GGGTTGGCCG	ATTCATTAAT	GCAGCTGGCA	6060
CGACAGGTTT	CCCGACTGGA	AAGCGGGCAG	TGAGCGCAAC	GCAATTAATG	TGAGTTAGCT	6120
CACTCATTAG	GCACCCAGG	CTTTACACTT	TATGCTTCGG	GCTCGTATGT	TGTGTGGAAT	6180
TGTGAGCGGA	TAACAATTTT	ACACAGGAAA	CAGCTATGAC	CAGGATGTAC	GAATTCGCAG	6240
GTAGGAGAGC	TGGCGGATC	CGAGGCTGAA	GGCGATGACC	CTGCTAAGGC	TGCATTCAAT	6300
AGTTTACAGG	CAAGTGCTAC	TGAGTACATT	GGCTACGCTT	GGGCTATGGT	AGTAGTTATA	6360
GTGCGTGCTA	CCATAGGGAT	TAAATTATTC	AAAAAGTTTA	CGAGCAAGGC	TTCTTAACCA	6420
GCTGGCGTAA	TAGCGAAGAG	GCCCGCACCG	ATCGCCCTTC	CCAACAGTTG	CGCAGCCTGA	6480
ATGGCGAATG	GCGCTTTGCC	TGGTTTCCGG	CACCAGAAGC	GGTGCCGGA	AGCTGGCTGG	6540
AGTGCGATCT	TCCTGAGGCC	GATACGGTGG	TGCTCCCCTC	AAACTGGCAG	ATGCACGGTT	6600
ACGATGCGCC	CATCTACACC	AACGTAACCT	ATCCCATTAC	GGTCAATCCG	CCGTTTGTTC	6660
CCACGGAGAA	TCCGACGGGT	TGTTACTCGC	TCACATTIAA	TGTTGATGAA	AGCTGGCTAC	6720
AGGAAGGCCA	GACGCGAATT	ATTTTGTATG	GCGTTCCTAT	TGGTTAAAAA	ATGAGCTGAT	6780
TTAACAAAAA	TTTAACGCGA	ATTTTAACAA	AATATTAACG	TTTACAATTT	AAATATTTGC	6840
TTATACAATC	TTCTGTTTTT	TGGGGCTTTT	CTGATTATCA	ACCGGGGTAC	ATATGATTGA	6900
CATGCTAGTT	TTACGATTAC	CGTTCATCGA	TTCTCTTGTT	TGCTCCAGAC	TCTCAGGCAA	6960
TGACCTGATA	GCCTTTGTAG	ATCTCTCAAA	AATAGCTACC	CTCTCCGGCA	TTAATTTATC	7020
AGCTAGAACG	GTGAATATC	ATATTGATGG	TGATTGACT	GTCTCCGGCC	TTTCTCACCC	7080
TTTTGAATCT	TTACCTACAC	ATTACTCAGG	CATTGCATT	AAAATATATG	AGGGTTCTAA	7140
AAATTTTTAT	CCTTGCGTTG	AAATAAAGGC	TTCTCCCGCA	AAAGTATTAC	AGGGTCATAA	7200
TGTTTTTGGT	ACAACCGATT	TAGCTTTATG	CTCTGAGGCT	TTATTGCTTA	ATTTTGCTAA	7260
TTCTTTGCCT	TGCTGTATG	ATTTATTGGA	CGTT			7294

(2) INFORMATION FOR SEQ ID NO:6:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7394 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: circular

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGCGCCCC AAATGAAAAAT

ATAGCTAAAC AGGTTATTGA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT 120
CGTTCCGAGA ATTGGGAATC AACTGTTAGA TGAATGAAA CTTCCAGACA CCGTACTTTA 180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA 240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG 300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TAAAAACGCG ATATTGAAG 360
TCTTTCCGGC TTCCTCTTAA TCTTTTIGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT 420
CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACT GTTTAAAGCA 480
TTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT 540
AAACATTTTA CTATTACCCC CTCIGGCAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT 600
GGTTTTATC GTCGTCTGGT AAACGAGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT 660
AATTCCTTTT GCGGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG 720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT 780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA 840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTC 900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG 960
AATATCCGGT TCTTGCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC 1020
TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CCGTTCCCTT ATGATTGACC 1080
GTCTGCGCCT CGTTCGGGCT AAGTAACATG GAGCAGGTG CCGATTTCGA CACAATTTAT 1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTCGCGC TTGGTATAAT CGGTGGGGGT 1200
CAAAGATGAG TGTTTTAGTG TATTCTTTCG CCTCTTTCGT TTIAGGTTGG TGCCTTCGTA 1260
GTGGCATTAC GTATTTTACC CGTTTAATGG AAACCTCCTC ATGAAAAAGT CTTIAGTCCT 1320
CAAAGCCTCT GTAGCCGTTG CTACCCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA 1380
CGATCCCGCA AAAGCGGCCT TTAACCTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA 1440
TGGGTGGGCG ATGGTTGTTG TCATTGTGGG CGCAACTATC GGIATCAAGC TGTTTAAGAA 1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTTT GGAGCCTTTT 1560
TTTTTGGAGA TTTTCAACGT GAAAAATTA TTATTCGCAA TTCCTTTAGT TGTTCCTTTC 1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTAGCAA AACCCCATAC AGAAAAATTCA 1680
TTTACTAACG TCTGGAAAGA CGACAAACT TIAGATCGTT ACGCTAACTA TGAGGGTTGT 1740
CTGTGGAATG CTACAGGCGT TGTAGTTTGT ACTGGTGACG AAACCTCAGTG TTACGGTACA 1800
TGGGTTCCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT 1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT 1920
ATTCCGGGCT ATACTTATAT CAACCTCTC GACGGCACTT ATCCGCCTGG TACTGAGCAA 1980
AACCCCGCTA ATCTAATCC TTCTCTTGAG GAGTCTCAGC CTCCTAATAC TTTCATGTTT 2040
CAGAATAATA GGTTCGAAA TAGGCAGGGG GCATTAACCTG TTTATACGGG CACTGTTACT 2100

CAAGGCACTG ACCCCGTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG 2160
TATGACGCTT ACTGGAACGG TAAATTCAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA 2220
GATCCATTCC TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCTGTCAAT 2280
GCTGGCGGGG GCTCTGGTGG TGGTCTGGT GCGGCTCTG AGGGTGGTGG CTCTGAGGGT 2340
GGCGGTTCTG AGGGTGGGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC TGGTTCGGT 2400
GATTTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGG CTATGACCGA AAATGCCGAT 2460
GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAAGTTGATT CTGTGCTAC TGATTACGGT 2520
GCTGCTATCG ATGGTTTCAT TGGTGACGT TCCGGCCTTG CTAATGGTAA TGGTGCTACT 2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT 2640
TTAATGAATA ATTTCCGTCA ATATTIACCT TCCCTCGCTC AATCGGTTGA ATGTGCCCCT 2700
TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTCTATTG ATTGTGACAA AATAAACTTA 2760
TTCCGTGGTG TCTTTGCGT TCTTTTATAT GTTGCCACCT TTATGTATGT ATTTCTIACG 2820
TTTGCTAACA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATTCCGT 2880
TATTATTGCG TTTCCTCGGT TTCCTTCTGG TAAGTTTGT CCGCTATCTG CTTACTTTTC 2940
TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTTTCATT GTTCTTGCT CTTATTATTG 3000
GGCTTAACTC AATTCTTGTG GGTATCTCT CTGATATTAG CGCTCAATTA CCCTCTGACT 3060
TTGTTCAAGG TGTTCACTTA ATTCTCCCGT CTAATGCGCT TCCCTGTTTT TATGTTATTC 3120
TCTCTGTAAA GGCTGCTATT TTCATTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG 3180
ATTGGGATAA ATAATATGGC TGTTTATTTT GTAAGTGGCA AATTAGGCTC TGGAAAGACG 3240
CTCGTTAGCG TTGTAAGAT TTAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT 3300
CTTGATTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAC GCCTCGCGTT 3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTGTCTG CTATTGGGCG CGGTAATGAT 3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGGGTAC TTGGTTAAT 3480
ACCGGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTTCT ACATGCTCGT 3540
AAATTAGGAT GGGATATTAT TTTCTTGTG CAGGACTTAT CTATTGTGA TAAACAGGCG 3600
CGTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCTGCTC TGGACAGAAT TACTTTACCT 3660
TTTGTCCGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT 3720
GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTGAGCG TTGGCTTTAT 3780
ACTGGTAAGA ATTTGIATAA CGCATATGAT ACTAAACAGG CTTTTTCTAG TAATTATGAT 3840
TCCGGTGTTT ATTCTTATT AACGCCTTAT TTATCACAGG GTCGGTATTT CAAACCATT 3900
AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTTGA AAAAGTTTTT ACGGTTCTT 3960
TGTCTTGGCA TTGGATTGTC ATCAGCATTT ACATATAGTT ATATAACCCA ACCTAAGCCG 4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCATAT TGACTCTTCT 4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAAATTAAT 4140

AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTTCC	4200
ATTAAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTTTGTTT TCTTGATGTT	4260
TGTTTCATCA TCTTCTTTTG CTCAGGTAAT TGAAATGAAT AATTCGCCTC TCGCGGATTT	4320
TGTAACCTGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTCTCCCG ATGTAAAAGG	4380
TACTGTTACT GTATATTCAT CTGACGTAA ACCTGAAAAT CTACGCAATT TCTTTATTTT	4440
TGTTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCCATAA TTCAGAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCTTCTG GTGGTTTCTT TGTTCCGCAA AATGATAATG TTAATCAAAC	4620
TTTTAAAATT AATAACGTTT GGGCAAAGGA TTAAATACGA GTTGTGCAAT TGTTTGTAAG	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT	4740
TAGTGCACCT AAAGATATTT TAGATAACCT TCCTCAATTC CTTTCTACTG TTGATTTGCC	4800
AACTGACCAG ATATTGATTG AGGGTTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA	4860
TTTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GCGCGTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTCGTTGGT ATTTTAAATG GCGATGTTTT	4980
AGGGCTATCA GTTCGCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTTAT	5100
TACTGGTGGT GTGACTGGTG AATCTGCCAA TGTAATAAAT CCATTTTACA CGATTGAGCG	5160
TCAAAATGTA GGTATTTCCA TGAGCGTTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT	5220
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT	5280
TACTAATCAA AGAAGTATTG CTACAACGGT TAATTTGCGT GATGGACAGA CTCTTTTACT	5340
CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GCGGTACCGT TCCTGTCTAA	5400
AATCCCTTTA ATCGGCCTCC TGTTTACGTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT	5460
ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCGCATTG AGCGCGGCGG	5520
GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT	5580
TCGCTTTCTT CCCTTCCTTT CTGCGCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC	5640
GGGGGCTCCC TTTAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAACTTG	5700
ATTTGGGTGA TGGTTACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA	5760
CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTTCCA AACTGGAACA AACTCAACC	5820
CTATCTCGGG CTATTCTTTT GATTTATAAG GGATTTTGCC GATTTCGGAA CCACCATCAA	5880
ACAGGATTTT CGCCTGCTGG GGCAAACCAG CGTGGACCGC TTGCTGCAAC TCTCTCAGGG	5940
CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCCT	6000
GGCGCCCAAT ACGCAAACCG CCTCTCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC	6060
ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAA GTGAGTTAGC	6120
TCACTGATTA GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA	6180

TTGTGAGCGG ATAACAATTT CACACGGCTC ACTTGGCACT GGCCGTCGTT TTACAACGTC	6240
GTGACTGGGA AAACCCTGGC GTTACCCAAG CTTTGACAT GGAGAAAATA AAGTGAAACA	6300
AAGCACTATT GCACTGGCAC TCTTACCGTT ACTGTTTACC CCTGTGGCAA AAGCCCTTCT	6360
GAGGCATCCG GGAGCTGAAG GCGATGACCC TGCTAAGGCT GCATTCAATA GTTTACAGGC	6420
AAGTGCTACT GAGTACATTG GCTACGCTTG GGCTATGGTA GTAGTTATAG TTGGTGCTAC	6480
CATAGGGATT AAATTATTCA AAAAGTTTAC GAGCAAGGCT TCTTAAGCAA TAGCGAAGAG	6540
GGCCGGACCG ATCGCCCTTC CCAACAGTTG CGCAGCCTGA ATGGCGAATG GCGCTTTGCC	6600
TGGTTTCCGG CACCAGAAGC GGTGCCGGAA AGCTGGCTGG AGTGCGATCT TCCTGAGGCC	6660
GATACGGTCG TCGTCCCCTC AAAGTGGCAG ATGCACGGTT ACGATGCGCC CATCTACACC	6720
AACGTAACCT ATCCCATTAC GGTCAATCCG CCGTTTGTTT CCACGGAGAA TCCGACGGGT	6780
TGTTACTCGC TCACATTTAA TGTGATGAA AGCTGGCTAC AGGAAGGCCA GACGCGAATT	6840
ATTTTGTATG GCGTTCCTAT TGGTTAAAAA ATGAGCTGAT TTAACAAAAA TTAAACGCCA	6900
TGGGGCTTTT CTGATTATCA ACCGGGGTAC ATATGATTGA CATGCTAGTT TTACGATTAC	6960
CGTTCATCGA TTCTCTTGTT TGCTCCAGAC TCTCAGGCAA TGACCTGATA GCCTTTGTAG	7020
ATCTCTCAAA AATAGCTACC CTCTCCGGCA TTAATTTATC AGCTAGAACG GTTGAATATC	7080
ATATTGATGG TGATTGACT GTCTCCGGCC TTTCTCACCC TTTTGAATCT TTACCTACAC	7140
ATTACTCAGG CATTGCATTT AAAATATATG AGGGTTCTAA AAATTTTAT CTTTGCGTTG	7200
AAATAAAGGC TTCTCCCGCA AAAGTATTAC AGGGTCATAA TGTTTTTGGT ACAACCGATT	7260
TAGCTTTATG CTCTGAGGCT TTATTGCTTA ATTTTGCTAA TTCTTTGCCT TGCCTGTATG	7320
ATTTATTGGA CGTT	7380
	7394

(2) INFORMATION FOR SEQ ID NO:7:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 37 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATCCTAGGC TGAAGGCGAT GACCCTGCTA AGGCTGC

(2) INFORMATION FOR SEQ ID NO:8:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

93

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATTCAATAGT TTACAGGCAA GTGCTACTGA GTACA

35

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TTGGCTACGC TTGGGCTATG GTAGTAGTTA TAGTT

35

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGTGCTACCA TAGGGATTAA ATTATTCAAA AAGTT

35

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TACGAGCAAG GCTTCTTA

18

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGCTTAAGAA GCCTTGCTCG TAAACTTTTT GAATAATTT

39

94

(2) INFORMATION FOR SEQ ID NO:13:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AATCCCTATG GTAGCACCAA CTATAACTAC TACCAT

36

(2) INFORMATION FOR SEQ ID NO:14:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGCCCAAGCG TAGCCAATGT ACTCAGTAGC ACTTG

35

(2) INFORMATION FOR SEQ ID NO:15:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCTGTAAACT ATTGAATGCA GCCTTAGCAG GCTC

34

(2) INFORMATION FOR SEQ ID NO:16:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ATCGCCTTCA GCCTAG

16

(2) INFORMATION FOR SEQ ID NO:17:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CTCGAATTCG TACATCCTGG TCATAGC

27

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CATTTTTCGA GATGGCTTAG A

21

(2) INFORMATION FOR SEQ ID NO:19:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TAGCATTAAAC GTCCAATA

18

(2) INFORMATION FOR SEQ ID NO:20:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATATATTTTA GTAAGCTTCA TCTTCT

26

(2) INFORMATION FOR SEQ ID NO:21:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GACAAAGAAC GCGTGAAAAC TTT

23

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCGGGCCTCT TCGCTATTGC TTAAGAAGCC TTGCT

35

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TTGAGCCTAG GATCGCCGA GCTCTCTAC CTGCGAATTC GTACATCC

48

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TGGATTATAC TTCTAAATAA TGGA

24

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TAACACTCAT TCCGGATGGA ATTCTGGAGT CTGGGT

36

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AATTCGCCAA GGAGACAGTC AT

22

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AATGAAATAC CTATTGCCCTA CGGCAGCCGC TGGATTGTT

39

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ATTACTCGCT GCCCAACCAG CCATGGCCGA GCTCGTGAT

39

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GACCCAGACT CCAGATATCC AACAGGAATG AGTGTTAAT

39

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TCTAGAACGC GTC

13

98

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

ACGTGACGCG TTCTAGAATT AACACTCATT CCTGT

35

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TGGATATCTG GAGTCTGGGT CATCAGGAGC TCGGCCATG

39

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GCTGGTTGGG CAGCGAGTAA TAACAATCCA GCGGCTGCC

39

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GTAGGCAATA GGTATTTTCAT TATGACTGTC CTTGGCC

37

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

TGACTGTCTC CTGGCGTGT GAAATTGTTA

30

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TAACACTCAT TCCGGATGGA ATTCTGGAGT CTGGGT

36

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CAATTTTATC CTAAATCTTA CCAAC

25

(2) INFORMATION FOR SEQ ID NO:38:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CATTTTIGCA GATGGCTTAG A

21

(2) INFORMATION FOR SEQ ID NO:39:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CGAAAGGGGG GTGTGCTGCA A

21

100

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TAGCATTAAAC GTCCAATA

18

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AAACGACGGC CAGTGCCAAG TGACGCGTGT GAAATTGTTA TCC

43

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GGCGAAAGGG AATTCTGCAA GGCGATTAAG CTTGGGTAAAC GCC

43

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGCGTTACCC AAGCTTTGTA CATGGAGAAA ATAAAG

36

101

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TGAAACAAAG CACTATTGCA CTGGCACTCT TACCGTTACC GT

42

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

TACTGTTTAC CCCTGTGACA AAAGCCGCCC AGGTCCAGCT GC

42

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TCGAGTCAGG CCTATTGTGC CCAGGGATTG TACTAGTGGA TCCG

44

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 38 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TGGCGAAAGG GAATTCGGAT CCACTAGTAC AATCCCTG

38

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

102

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GGCACAATAG GCCTGACTCG AGCAGCTGGA CCAGGGCGGC TT

42

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

TTGTCACAGG GGTAACAGT AACGGTAACG GTAAGTGTGC CA

42

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GTGCAATAGT GCTTTGTTTC ACTTTATTTT CTCCATGTAC AA

42

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TAACGGTAAG AGTGCCAGTG C

21

(52) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: misc difference
- (B) LOCATION: replace(25, "")
- (D) OTHER INFORMATION: /note= "M REPRESENTS AN EQUAL MIXTURE OF A AND C AT THIS LOCATION AND AT LOCATIONS 28, 31, 34, 37, 40, 43, 46 & 49"

103

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

AGCTCCCGGA TGCCTCAGAA GATGMNNMNN MNNMNNMNNM NNMNNMNNMN NGGCTTTTGC .60
CACAGGGG .68

(2) INFORMATION FOR SEQ ID NO:53:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: misc difference
(B) LOCATION: replace(17, "")
(D) OTHER INFORMATION: /note="M REPRESENTS AN EQUAL
MIXTURE OF A AND C AT THIS LOCATION AND AT
LOCATIONS 20, 23, 26, 29, 32, 35, 38, 41, 44 & 50"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CAGCCTCGGA TCCGCCMNNM NNMNNMNNMN NMNNMNNMNN MNNMNNATCM GAAT .54

(2) INFORMATION FOR SEQ ID NO:54:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GGTAAACAGT AACGGTAAGA GTGCCAG .27

(2) INFORMATION FOR SEQ ID NO:55:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GGGCTTTTGC CACAGGGGT .19

(2) INFORMATION FOR SEQ ID NO:56:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

104

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:56:

AGGGTCATCG CCTTCAGCTC CGGATCCCTC AGAAGTCATA AACCCCCCAT AGGCTTTTGC

60

CAC

63

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 47 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:57:

TCGCCTTCAG CTCCCGGATG CCTCAGAAGC ATGAACCCCC CATAGGC

47

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CAATTTTATC CTAAATCTTA CCAAC

25

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GCCTTCAGCC TCGGATCCGC C

21

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CGGATGCCTC AGAAGGCCCN N

21

105

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:61:

CGGATGCCTC AGAAGGGCTT TTGCCACAGG

30

I CLAIM:

1. A composition of matter comprising a plurality of cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, said expressible oligonucleotides having a desirable bias of random codon sequences produced from random combinations of first and second oligonucleotide precursor populations having a desirable bias of random codon sequences.

2. The composition of claim 1, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

3. The composition of claim 1, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is biased toward a predetermined sequence.

4. The composition of claim 1, wherein said first and second oligonucleotides having random codon sequences have at least one specified codon at a predetermined position.

5. The composition of claim 1, wherein said cells are procaryotes.

6. The composition of claim 1, wherein said cells are E. coli.

7. A kit for the preparation of vectors useful for the expression of a diverse population of random peptides from combined first and second oligonucleotides having a desirable bias of random codon sequences, comprising: two vectors: a first vector having a cloning site for said first oligonucleotides and a pair of restriction sites for operationally combining first oligonucleotides with second oligonucleotides; and a second vector having a cloning site for said second oligonucleotides and a pair of restriction sites complementary to those on said first vector, one or both vectors containing expression elements capable of being operationally linked to said combined first and second oligonucleotides.

8. The kit of claim 7, wherein said vectors are in a filamentous bacteriophage.

9. The kit of claim 8, wherein said filamentous bacteriophage are M13.

10. The kit of claim 7, wherein said vectors are plasmids.

11. The kit of claim 7, wherein said vectors are phagemids.

12. The kit of claim 7, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

13. The kit of claim 7, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

14. The kit of claim 7, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

15. The kit of claim 7, wherein said pair of restriction sites are Fok I.

16. A cloning system for expressing random peptides from diverse populations of combined first and second oligonucleotides having a desirable bias of random codon sequences, comprising: a set of first vectors
5 having a diverse population of first oligonucleotides having a desirable bias of random codon sequences and a set of second vectors having a diverse population of second oligonucleotides having a desirable bias of random codon sequences, said first and second vectors each
10 having a pair of restriction sites so as to allow the operational combination of first and second oligonucleotides into a contiguous oligonucleotide having a desirable bias of random codon sequences.

17. The cloning system of claim 16, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

18. The cloning system of claim 16, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

19. The cloning system of claim 16, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

20. The cloning system of claim 16, wherein said combined first and second vectors is through a pair of restriction sites.

21. The cloning system of claim 16, wherein said pair of restriction sites are Fok I.

22. A composition of matter comprising a plurality of cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, said expressible oligonucleotides
5 having a desirable bias of random codon sequences.

23. The composition of claim 22, wherein said cells are procaryotes.

24. The composition of claim 22, wherein said expressible oligonucleotides are expressed as peptide fusion proteins on the surface of a filamentous bacteriophage.

25. The composition of claim 22, wherein said filamentous bacteriophage is M13.

26. The composition of claim 22, wherein said fusion protein contains the product of gene VIII.

27. The composition of claim 22, wherein said diverse population of oligonucleotides having a desirable bias of random codon sequences are produced from the combination of diverse populations of first and second
5 oligonucleotides having a desirable bias of random codon sequences.

28. The composition of claim 22, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.

29. The composition of claim 22, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

30. The composition of claim 22, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

31. A plurality of vectors containing a diverse population of expressible oligonucleotides having a desirable bias of random codon sequences.

32. The vectors of claim 31, wherein said oligonucleotides are expressible as fusion proteins on the surface of filamentous bacteriophage.

33. The vectors of claim 31, wherein said filamentous bacteriophage is M13.

34. The vectors of claim 31, wherein said fusion protein contains the product of gene VIII.

35. The vectors of claim 31, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.

36. The vectors of claim 31, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

37. The vectors of claim 31, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

38. A composition of matter, comprising a diverse population of oligonucleotides having a desirable bias of random codon sequences produced from random combinations of two or more oligonucleotide precursor
5 populations having a desirable bias of random codon sequences.

39. A method of constructing a diverse population of vectors having combined first and second oligonucleotides having a desirable bias of random codon sequences capable of expressing said combined
5 oligonucleotides as random peptides, comprising the steps of:

- 10 (a) operationally linking sequences from a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;
- 15 (b) operationally linking sequences from a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and
- 20 (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors capable of being expressed.

40. The method of claim 39, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

41. The method of claim 39, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

42. The method of claim 39, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

43. The method of claim 38, wherein steps (a) through (c) are repeated two or more times.

44. A method of selecting a peptide capable of being bound by a ligand binding protein from a population of random peptides, comprising:

- 5 (a) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;
- 10 (b) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector;
- 15 (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors;
- 20 (d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of random peptides; and
- (e) determining the peptide which binds to said ligand binding protein.

45. The method of claim 44, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

46. The method of claim 44, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

47. The method of claim 44, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

48. The method of claim 44, wherein steps (a) through (c) are repeated two or more times.

49. A method for determining the nucleic acid sequence encoding a peptide capable of being bound by a ligand binding protein which is selected from a population of random peptides, comprising:

- 5 (a) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;
- 10 (b) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector;
- 15 (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors;
- 20 (d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of random peptides;
- (e) determining the peptide which binds to said ligand binding protein;
- 25 (f) isolating the nucleic acid encoding said peptide; and
- (g) sequencing said nucleic acid.

50. The method of claim 49, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

51. The method of claim 49, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

52. The method of claim 49, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

53. The method of claim 49, wherein steps (a) through (c) are repeated two or more times.

54. A method of constructing a diverse population of vectors containing expressible oligonucleotides having a desirable bias of random codon sequences, comprising operationally linking a diverse
5 population of oligonucleotides having a desirable bias of random codon sequences to expression elements.

55. The method of claim 54, wherein said oligonucleotides are expressible as fusion proteins on the surface of filamentous bacteriophage.

56. The method of claim 54, wherein said filamentous bacteriophage are M13.

57. The method of claim 54, wherein said fusion protein contains the product of gene VIII.

58. The method of claim 54, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.

59. The method of claim 54, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

60. The method of claim 54, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

61. The method of claim 54, wherein said operationally linking further comprising the steps of:

5 (a) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;

10 (b) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and

15 (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors.

62. The method of claim 61, wherein steps (a) through (c) are repeated two or more times.

63. A method of selecting a peptide capable of being bound by a binding protein from a population of random peptides, comprising:

- 5 (a) operationally linking a diverse population of oligonucleotides having a desirable bias of random codon sequences to expression elements;
- 10 (b) introducing said population of vectors into a compatible host under conditions sufficient for expressing said population of random peptides; and
- (c) determining the peptide which binds to said ligand binding protein.

64. The method of claim 63, wherein said population of random peptides are expressed as fusion proteins on the surface of filamentous bacteriophage.

65. The method of claim 63, wherein said filamentous bacteriophage are M13.

66. The method of claim 63, wherein said fusion protein contains the product of gene VIII.

67. The method of claim 63, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.

68. The method of claim 63, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

69. The method of claim 63, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

70. The method of claim 63, wherein step (a) further comprises:

5 (a1) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;

10 (a2) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and

15 (a3) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors.

71. The method of claim 70, wherein steps (a1) through (a3) are repeated two or more times.

72. A method of determining the nucleic acid sequence encoding a peptide capable of being bound by a ligand binding protein which is selected from a population of random peptides, comprising:

- 5 (a) operationally linking a diverse population of oligonucleotides having a desirable bias of random codon sequences to expression elements.
- 10 (b) introducing said population of vectors into a compatible host under conditions sufficient for expressing said population of random peptides;
- (c) determining the peptide which binds to said ligand binding protein;
- 15 (d) isolating the nucleic acid encoding said peptide; and
- (e) sequencing said nucleic acid.

73. The method of claim 72, wherein said population of random peptides are expressed as fusion proteins on the surface of filamentous bacteriophage.

74. The method of claim 72, wherein said filamentous bacteriophage are M13.

75. The method of claim 72, wherein said fusion protein contains the product of gene VIII.

76. The method of claim 72, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.

77. The method of claim 72, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

78. The method of claim 72, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

79. The method of claim 72, wherein step (a) further comprises:

(a1) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;

(a2) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and

(a3) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors.

80. The method of claim 78, wherein steps (a1) through (a3) are repeated two or more times.

81. A vector comprising two copies of a gene encoding a filamentous bacteriophage coat protein, both copies encoding substantially the same amino acid sequence but having different nucleotide sequences.

82. The vector of claim 81, wherein said filamentous bacteriophage is M13.

83. The vector of claim 81, wherein said gene is gene VIII.

84. The vector of claim 81, wherein said vector has substantially the sequence shown in Figure 5 (SEQ ID NO: 1).

85. A vector comprising two copies of a gene encoding a filamentous bacteriophage coat protein, one copy of said gene capable of being operationally linked to an oligonucleotide wherein said oligonucleotide can be expressed as a fusion protein on the surface of said filamentous bacteriophage or as a soluble peptide.

86. The vector of claim 84, wherein said one copy of said gene is expressed on the surface of said filamentous bacteriophage.

87. The vector of claim 84, wherein said bacteriophage coat protein is M13 gene VIII.

1 / 16

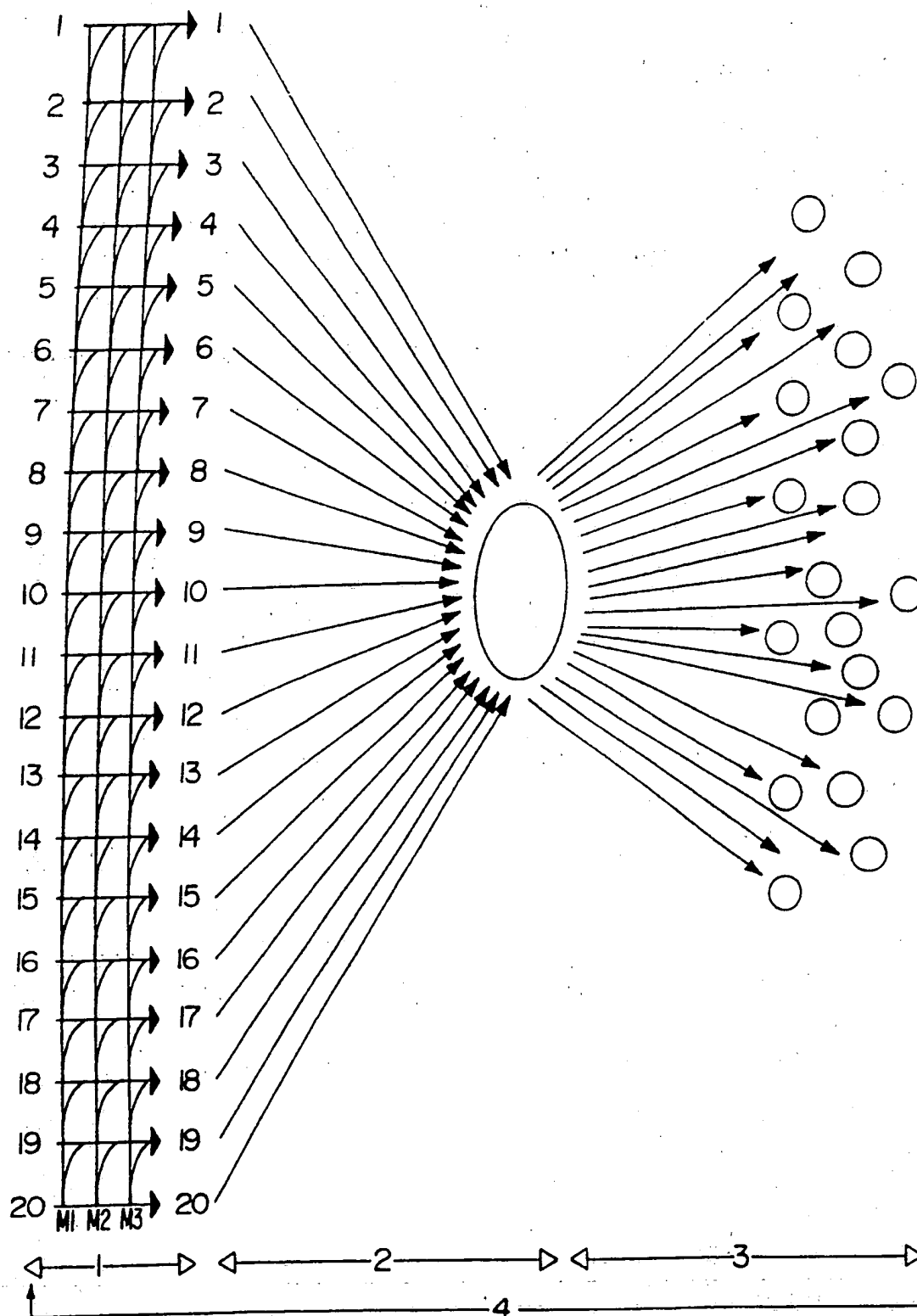


FIG. 1

2 / 16

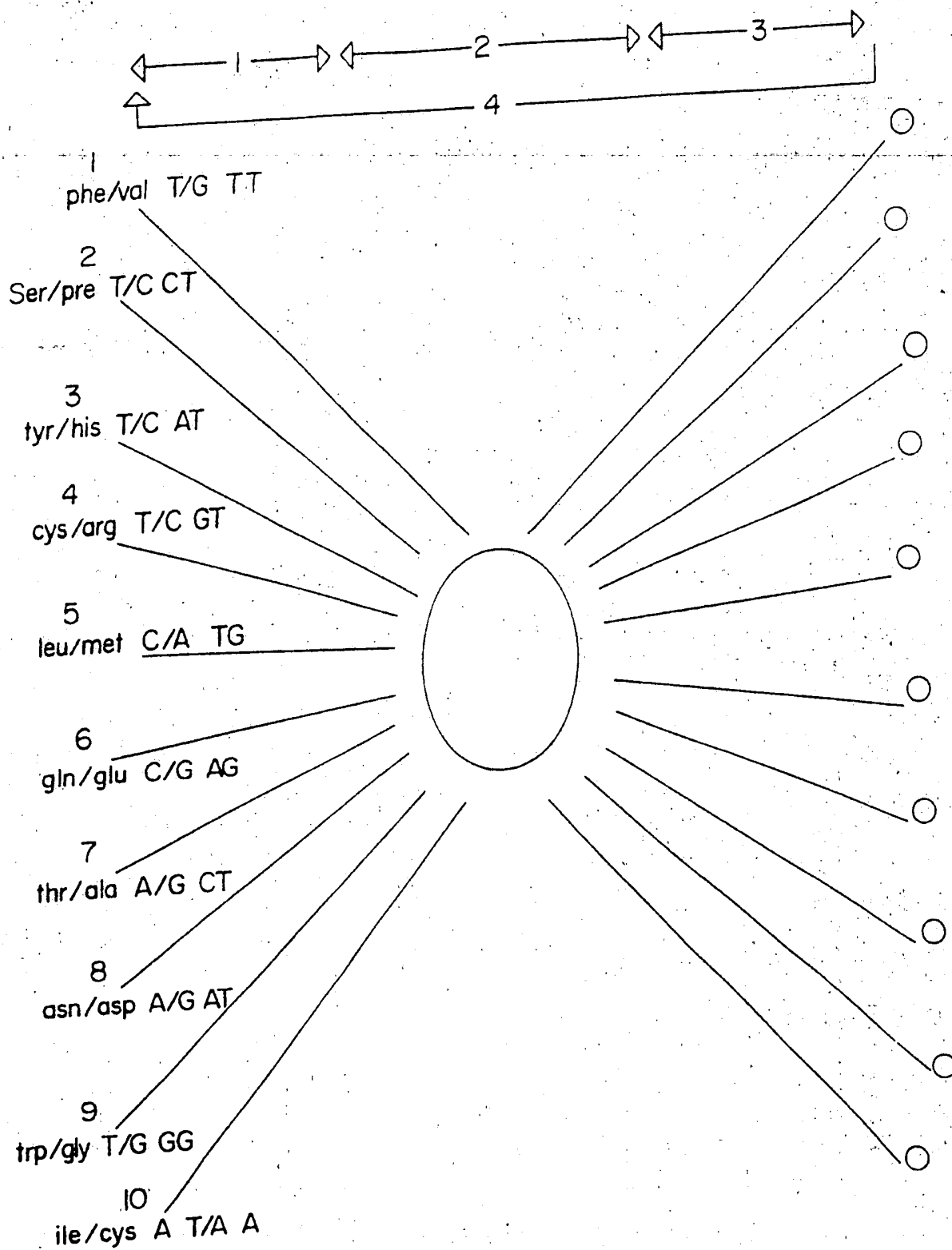
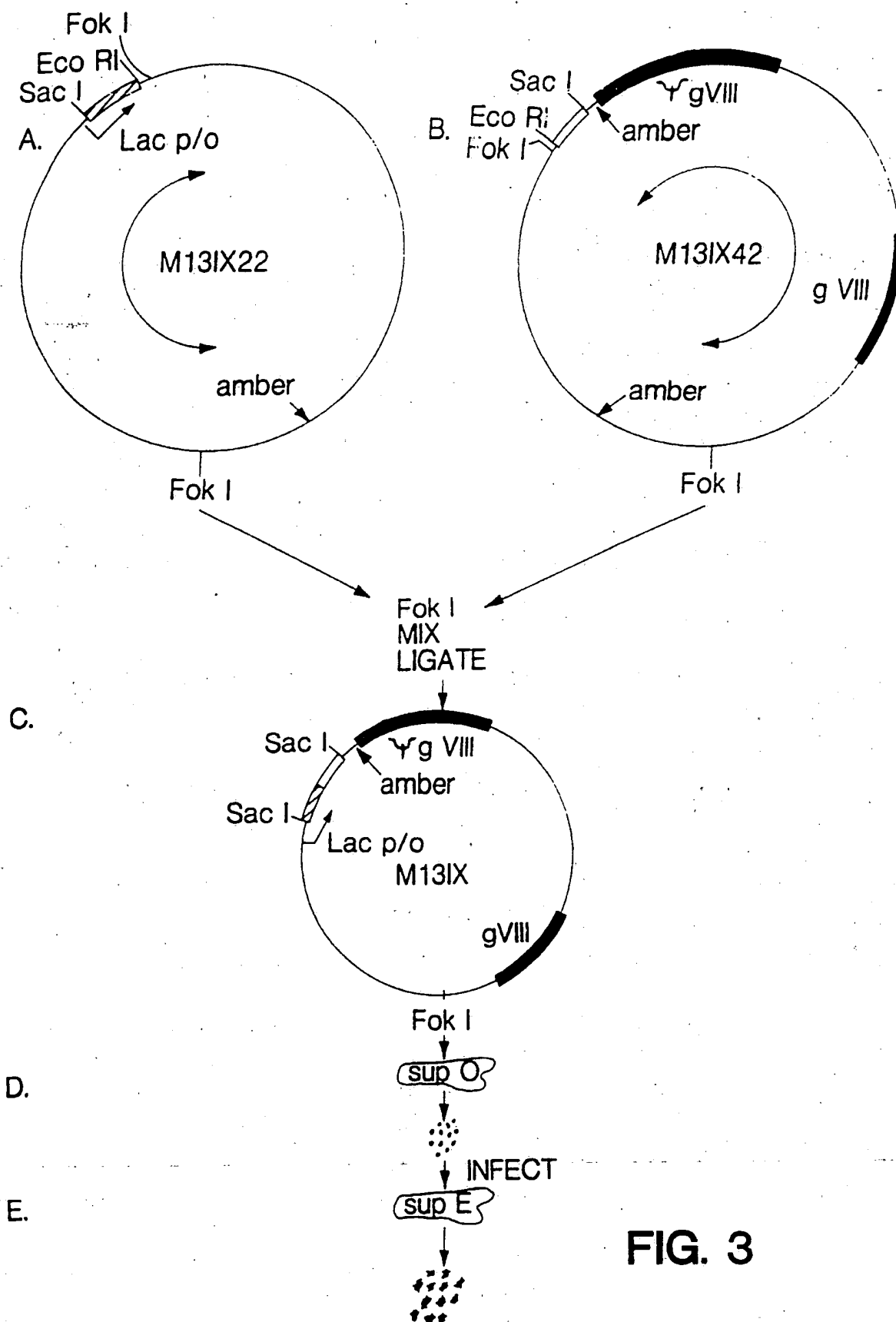


FIG. 2

3 / 16



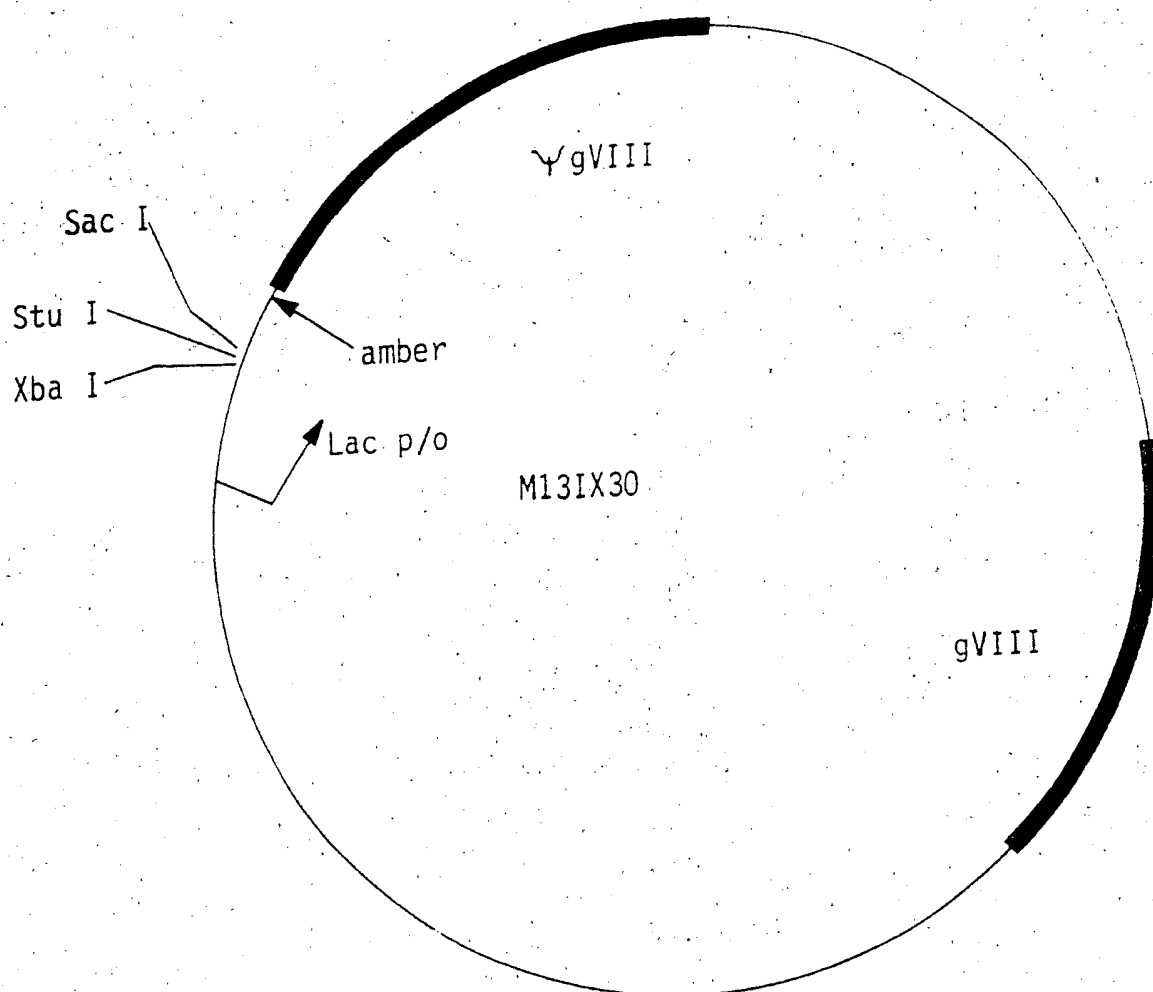


FIG. 4

5 / 16

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAAT
61	ATAGCTAAAC	AGGTTATTGA	CCATTTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT
121	CGTTCCGAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAGATTC	AGCAATTAAG	CTCTAAGCCA
241	TCTGCCAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAGG	IACCTCTAA	TCCTGACCTG
301	TTGGAGTTTG	CTTCCGGTCT	GGTTTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTTGAAG
361	TCTTTCGGGC	TTCTCTTAA	TCTTTTTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCGAG	TATTGGACGC	TATCCAGTCT
541	AAACATTTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTTG	CAAAAGCCTC	TCGCTATTTT
601	GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCTCGT
661	AATTCCTTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCCTAA	ATCTCAACTG
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATTTT
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCCTA	AAATCGCATA	AGGTAATTTA
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGTTC
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTIG	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC
1021	TGTACACCGT	TCATCTGTCT	TCTTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC
1081	GTCTGCGCCT	CGTTCCGGCT	AAGTAAACATG	GAGCAGGTCTG	CGGATTTCTGA	CACAATTTAT
1141	CAGGCGATGA	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT
1201	CAAAGATGAG	TGTTTTAGTG	TATTCCTTCG	CCTCTTTCGT	TTTAGGTTGG	TGCCTTCGTA
1261	GTGGCATTAC	GTTTTTACC	CGTTTAATGG	AAACTTCCTC	ATGAAAAAGT	CTTTAGTCCT
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCTCGT	TCCGATGCTG	TCTTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCGGCCT	TAACTCCCT	GCAAGCCTCA	CGGACCGAAT	ATATCGGTTA
1441	TGCGTGGGCG	ATGGTTGTTG	TCATTGTCTG	CGCAACTATC	GGTATCAAGC	TGTTTAAAGA
1501	ATTACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAG	GGCTCCTTTT	GGAGCCTTTT
1561	TTTTTGGAGA	TTTTCAACGT	GAAAAAATTA	TTATTCGCAA	TTCTTTTAGT	TGTTCTTTTC
1621	TATTCCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAAATTCA
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAAGT	TTAGATCGTT	ACGCTAATA	TGAGGGTTGT
1741	CTGTGGAATG	CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TACGGTACAT
1801	TGGGTTCTTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT
1921	ATTCCGGGCT	ATACCTATAT	CAACCCTCTC	GACGGCACTT	ATCCGCTTGG	TACTGAGCAA
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT
2041	CAGAATAACT	GGTTCGGA	TAGGCAAGGG	GCATTAAGTG	TTTATACGGG	CACTGTTACT
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAAGCCATG
2161	TATGACGCTT	ACTGGAACGG	TAAATTCAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA
2221	GATCCATTCTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TCGGTCAACC	TCCTGTCAAT
2281	GCTGGCGGCG	GCTCTGGTGG	TGGTCTGTGG	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTCCGGT
2401	GATTTTGATT	ATGAAAAGAT	GGCAACCGCT	AAATAAGGGG	CTATGACCGA	AAATGCCGAT
2461	GAAAACGCGC	TACAGTCTGA	CGCTAAGGCG	AAACTTGATT	CTGTGCGTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT
2581	GGTGATTTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAETCG	GTGACGGTGA	TAATTCACCT
2641	TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTTTCTATTG	ATTGTGACAA	AATAAACTTA
2761	TTCCGTGGTG	TCTTTGCGTT	TCTTTTATAT	TTTGCCACCT	TTATGTATGT	ATTTTCTACG
2821	TTTGCTAACA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTTG	GGTATTCGGT
2881	TATTATTGCG	TTTCCTCGGT	TTCTTCTGG	TAACCTTGTG	GCCGTATCTG	CTTACTTTTC
2941	TTAAAAGGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTTCAAT	GTTTCTTGCT	CTTATTATTG
3001	GGCTTAACCT	AATTCCTGTG	GGTTATCTCT	CTGATATTAG	CGCTCAATTA	CCCTCTGACT
3061	TTGTTCAAGG	GGTTCAAGTA	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTTT	TATGTTATTC
3121	TCTCTGTAAA	TGCTGCTATT	TTTATTTTGG	ACGTTAAACA	AAAAATCGTT	TCTTATTTGG
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAACCTGGC	AATTAGGCTC	TGGAAAGACG
3241	CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT
3301	CCTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAAC	GCCTCGCGTT
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGGCG	CGGTAATGAT
3421	TCCTACGATG	AAAATAAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGCGGTAC	TTGGTTTAAAT
3481	ACCCGTTCTT	GGAAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTTCT	ACATGCTCGT
3541	AAATTAGGAT	GGGATATTAT	CTTCCTTGTT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG
3601	CGTTCTGCAT	TAGCTGAACA	TCTTGTATAT	TGTCGTCGTC	TGGACAGAAT	TACTTTACCT
3661	TTTGTGCGTA	CTTTATATTC	TCTTATTACT	GGCTCGAAAA	TGCCTCTGCC	TAAATTACAT
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT
3781	ACTGGTAAGA	ATTTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT

FIG. 5-1

SUBSTITUTE SHEET

3841	TCCGGTGT	ATTCTTAT	AACGCCTT	TTATCACAC	GTCGGTAT	CAAACCAT	3900
3901	AATTTAGG	AGAAGATG	GCTTACTA	ATATATTT	AAAAGTTT	ACGCGTTC	3960
3961	TGTCTTGC	TTGGATTT	ATCAGCAT	ACATATAG	ATATAACCC	ACCTAAGCC	4020
4021	GAGGTTAA	AGGTAGTC	TCAGACCT	GATTTTGAT	AATTCACAT	TGACTCTTC	4080
4081	CAGCGTCT	ATCTAAGCT	TCGCTATG	TTCAAGGAT	CTAAGGGAA	ATTAATTA	4140
4141	AGCGACGA	TACAGAAG	AGGTTATTC	CTCACATAT	TTGATTTAT	TACTGTTTC	4200
4201	ATTA AAAAG	TAATTC AA	GAAATTTG	AATGTAAT	ATTTTGTTT	CTTGATGTT	4260
4261	GTTTCATCA	CTTCTTTG	TCAGGTAAT	GAAATGAAT	ATTCGCCTC	GCGCGATTT	4320
4321	GTAAC TTGG	ATTCAAAGC	ATCAGGCGA	TCCGTTATT	TTTCTCCCG	TGTAAAAGG	4380
4381	ACTGTTACT	TATATTCAT	TGACGTTAA	CCTGAAATC	TACGCAATT	CTTTATTTCT	4440
4441	GTTTACGTG	CTAATAATT	TGATATGGT	GGTTCAATT	CTTCCATTAT	TTAGAAGTAT	4500
4501	AATCCAAACA	ATCAGGATTA	TATTGATGA	TTGCCATCAT	CTGATAATCA	GGAATATGAT	4560
4561	GATAATTCG	CTCCTTCTG	TGGTTTCTT	GTTCCGCAAA	ATGATAATGT	TACTCAAAC	4620
4621	TTTAA AATTA	ATAACGTTG	GGCAAAGGAT	TTAATACGAG	TTGTCGAAT	GTTTGTAAG	4680
4681	TCTAATCTT	CTAAATCCTC	AAATGTATTA	TCTATTGACG	GCTCTAATCT	ATTAGTTGT	4740
4741	AGTGCACCTA	AAGATATTTT	AGATAACCTT	CCTCAATTCC	TTTCTACTGT	TGATTTGCCA	4800
4801	ACTGACCAGA	TATTGATTGA	GGGTTTGATA	TTTGAGGTTT	AGCAAGGTGA	TGCTTTAGAT	4860
4861	TTTTCA TTG	CTGCTGGCTC	TCAGCGTGCG	ACTGTTGCAG	GCGGTGTTAA	TACTGACCGC	4920
4921	CTCACCTCTG	TTTTATCTTC	TGCTGGTGGT	TCGTTCCGGTA	TTTTTAATGG	CGATGTTTTA	4980
4981	GGGCTATCAG	TTTCAGGTC	AAAGACTAAT	AGCCATTCAA	AAATATTGTC	TGTGCCACGT	5040
5041	ATTCCTACGC	TGACTGGTGA	GAAGGGTTCT	ATCTCTGTTG	GCCAGAAATGT	CCCTTTTATT	5100
5101	ACTGGTCGTG	GTATTTCCAT	ATCTGCCAAT	GTAATAATC	CATTTAGAC	GATTGAGCGT	5160
5161	CAAAATGTAG	CCAGCAAGGC	GAGCGTTTTT	CCTGTTGCAA	TGGCTGGCGG	TAATATTGTT	5220
5221	CTGGATATTA	GAAGTATTGC	CGATAGTTTG	AGTTCTTCTA	CTCAGGCAAG	TGATGTTATT	5280
5281	ACTAATCAAA	CTGATTATAA	TACAACGGTT	AATTTGCGTG	ATGGACAGAC	CTTTTACTC	5340
5341	GGTGGCTCA	TCGGCCTCCT	AAACACTTCT	CAAGATTCTG	GCGTACCGTT	CCTGTCTAAA	5400
5401	ATCCCTTTAA	TCAAAGCAAC	GTTTAGCTCC	CGCTCTGATT	CCAACGAGGA	AAGCACGTTA	5460
5461	TACGTGCTCG	ACGCGCAGCG	CATAGTACGC	GCCCTGTAGC	GGCGCATTA	GCGCGGCGGG	5520
5521	TGTGGTGGTT	CCTTCTTTTC	TGACCGCTAC	ACTTGCCAGC	GCCCTAGCGC	CCGCTCCTT	5580
5581	CGCTTTCTTC	TTAGGGTTCC	TCGCCACGTT	CGCCGGCTTT	CCCGGTCAAG	CTCTAAATCG	5640
5641	GGGGCTCCCT	GGTTACGTA	GATTTAGTGC	TTTACGGCAC	CCCGTCAAG	AAAACTTGA	5700
5701	TTTGGGTGAT	ACGTTCTTTA	GTGGGGCATC	CCCTGTATAG	ACTGGAACAA	GCCCTTTGAC	5760
5761	GTTGGAGTCC	TATTCTTTTG	ATAGTGGACT	CTTGTTCCTA	ATTTTCGGAAC	CACTCAACCC	5820
5821	TATCTCGGGC	GCCTGCTGGG	ATTTATAAGG	GATTTTGCCG	TGCTGCAACT	CACCATCAAA	5880
5881	CAGGATTTTC	AGGGCAATCA	GCAAACAGC	GTGGACCGCT	TGAAAAGAAA	CTCTCAGGGC	5940
5941	CAGGCGGTGA	CGCAAACCGC	GCTGTTGCC	GTCTCGCTGG	ATTCATTAAT	AACCACCCTG	6000
6001	GCGCCCAATA	CCCGACTGGA	CTCTCCCCGC	GCGTTGGCCG	GCAATTAATG	GCAGCTGGCA	6060
6061	CGACAGGTTT	GCACCCAGG	AAGCGGGCAG	TGAGCGCAAC	GCTCGTATGT	TGAGTTAGCT	6120
6121	CACCTATTAG	TAACAAATTC	CTTTACACTT	TATGCTTCCG	GCTCGTATGT	TGTGTGGAAT	6180
6181	TGTGAGCGGA	TCGGCGGATC	ACACAGGAAA	CAGCTATGAC	CAGGATGTAC	GAATTCGCAG	6240
6241	GTAGGAGAGC	CAAGTGCTAC	CTAGGCTGAA	GGCGATGACC	CTGCTAAGGC	TGCATTCAAT	6300
6301	AGTTTACAGG	CCATAGGGAT	TGAGTACATT	GGCTACGCTT	GGGCTATGGT	AGTAGTTATA	6360
6361	GTTGGTGCTA	TAGCGAAGAG	TAAATTATTC	AAAAAGTTTA	CGAGCAAGGC	TTCTTAACCA	6420
6421	GCTGGCGTAA	GCGCTTTGCC	GCCCGCACCG	ATCGCCCTTC	CCAACGATTG	CGCAGCCTGA	6480
6481	ATGGCGAATG	TCCTGAGGCC	TGGTTTCCGG	CACGAAAGC	GGTGCCGGAA	AGCTGGCTGG	6540
6541	AGTGCGATCT	CATCTACACC	GATACGGTGC	TCGTCCCTC	AAACTGGCAG	ATGCACGGTT	6600
6601	ACGATGCGCC	TGTTACTCGC	AACGTAACCT	ATCCCATTAC	GGTCAATCCG	CCGTTTGTTT	6660
6661	CCACGGAGAA	TCCGACGGGT	ATTTTGTATG	TCACATTTAA	TGTTGATGAA	AGCTGGCTAC	6720
6721	AGGAAGGCCA	GACGCGAATT	ATTTTATGAT	GCGTTTCTAT	TGGTTAAAAA	ATGAGCTGAT	6780
6781	TTAACA AAAA	TTTAACGCGA	ATTTTAAACA	AAATTAACG	TTTACAATTT	AAATATTTGC	6840
6841	TTATACAATC	TTCTGTTTTT	TGGGGCTTTT	CTGATTATCA	ACCGGGGTAC	ATATGATTGA	6900
6901	CATGCTAGTT	TTACGATTAC	CGTTGCTCGA	TTCTCTTGTT	TGCTCCAGAC	TCTCAGGCAA	6960
6961	TGACCTGATA	GCCTTTGTAG	ATCTCTCAAA	AATAGCTACC	CTCTCCGGCA	TTAATTTATC	7020
7021	AGCTAGAACG	GTTGAATATC	ATATTGATGG	TGATTTGACT	GTCTCCGGCC	TTTCTCACC	7080
7081	TTTTGAATCT	TTACCTACAC	ATTACTCAGG	CATTGCAATT	AAAATATATG	AGGGTTCTAA	7140
7141	AAATTTTAT	CCTTGCGTTG	AAATAAAGGC	TTCTCCCGCA	AAAGTATTAC	AGGGTCATAA	7200
7201	TGTTTTTGGT	ACAACCGATT	TAGCTTTATG	CTCTGAGGCT	TTATTGCTTA	ATTTTGCTAA	7260
7261	TTCTTTGCC	TGCCTGTATG	ATTTTATGGA	CGTT			7294

FIG. 5-2

SUBSTITUTE SHEET

7 / 16

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAT 60
61	ATAGCTAAAC	AGGTTATTGA	CCATTTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT 120
121	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA 180
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAGATTC	AGCAATTAAG	CTCTAAGCCA 240
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAGG	TACTCTCTAA	TCCTGACCTG 300
301	TTGGAGTTTG	CTTCCGGTCT	GGTTGCGTTT	GAAGCTCGAA	TTAAAAACGCG	ATATTGGAAG 360
361	TCTTTGCGGC	TTCTCTTAA	TCTTTTGTAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT 420
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA 480
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT 540
541	AAACATTTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTTG	CAAAAGCCTC	TCGCTATTTT 600
601	GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCCTGT 660
661	AATTCCTTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCCTAA	ATCTCAACTG 720
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATTTT 780
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA 840
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGTTC 900
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG 960
961	AAATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCCGCTGAGT 1020
1021	TGTACACCGT	TCATCTGTCC	TCTTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC 1080
1081	GTCTGCGCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTCG	CGGATTTCTGA	CACAATTTAT 1140
1141	CAGGCGATGA	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT 1200
1201	CAAAGATGAG	TGTTTTAGTG	TATCTTTTCG	CCTCTTTTCG	TTTAGGTTGG	TGCCTTCGTA 1260
1261	GTGGCATTAC	GTATTTTACC	CGTTTAAATG	AAACTTCCTC	ATGAAAAAGT	CTTTAGTCCT 1320
1321	CAAGCCCTCT	GTAGCCGTTG	CTACCCCTCG	TCCGATGCTG	TCTTTTCGCTG	CTGAGGGTGA 1380
1381	CGATCCCGCA	AAAGCGGCCT	TAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA 1440
1441	TGCGTGGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAAGAA 1500
1501	ATTACACCTG	AAAGCAAGCT	GATAAACCGA	TACAATTAAT	GGCTCCTTTT	GGAGCCTTTT 1560
1561	TTTTTGAGAA	TTTTCAACGT	GAAAAAATTA	TTATTGCGAA	TTCTTTTAGT	TGTTCTTTTC 1620
1621	TATTTCTACT	CCGTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AAACCCATAC	AGAAAAATTA 1680
1681	TTTTACTAAG	TCTGAAAGAA	CGACAAACTT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT 1740
1741	CTGTGGAATG	CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA 1800
1801	TGGGTTCTTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT 1860
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT 1920
1921	ATTCGGGGCT	ATACTTATAT	CAACCCCTCT	GACGGCACTT	ATCCGCCTGG	TACTGAGCAA 1980
1981	AACCCCGCTA	ACTCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT 2040
2041	CAGAATAATA	GGTCCGAAA	TAGGCAGGGG	GCAATTAATG	TTTATACGGG	CACCTGTTACT 2100
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG 2160
2161	TATGACGCTT	ACTGGAACGG	TAAATTCAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA 2220
2221	GATCCATTCT	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCTCAACC	TCCTGTCAAT 2280
2281	GCTGGCGGCG	GCTCTGGTGG	TGGTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT 2340
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTTCGGT 2400
2401	GATTTTGATT	ATGAAAAGAT	GGCAACGGG	AATAAGGGG	CTATGACCGA	AAATGCCGAT 2460
2461	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT 2520
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT 2580
2581	GGTGATTTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCACCT 2640
2641	TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT 2700
2701	TTTGCTTTTA	GCGCTGGTAA	ACCATATGAA	TTTTCTATTG	ATTGTGACAA	AATAAACTTA 2760
2761	TTCCGTGGTG	TCTTTGCGTT	TCTTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTTCTACG 2820
2821	TTTGCTAACA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCCTTTG	GGTATTCCTG 2880
2881	TATTATTGCG	TTTCTCGGT	TTCTTCTGG	TAACCTTGTG	CGGCTATCTG	CTTACTTTTC 2940
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTTATT	GTTTCTTGCT	CTTATTATTG 3000
3001	GGCTTAACTC	AATTTCTTGT	GGTTATCTCT	CTGATATTAG	CGCTCAATTA	CCCTCTGACT 3060
3061	TGTTTCAAGG	TGTTCAAGTTA	ATTCTCCCGT	GTAATGCGCT	TCCCTGTTTT	TATGTTATT 3120
3121	TCTCTGTAAA	GGCTGCTATT	TTCAATTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTTGG 3180
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAAGTGGCA	AATTAGGCTC	TGGAAAGACG 3240
3241	CTCGTTAGCG	TTGGTAAGAT	TAGGATAAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT 3300
3301	CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAAAC	GCCTCGCGTT 3360
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGGCG	CGGTAATGAT 3420
3421	TCTACAGATG	AAAAATAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGCGGTAC	TTGGTTTAAAT 3480
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTTCT	ACATGCTCGT 3540
3541	AAATTAGGAT	GGGATATTAT	CTTCTTGTG	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG 3600
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTTAT	TGTCGTCGTC	TGGACAGAAT	TACTTTACCT 3660
3661	TTTGTGCGTA	CTTTATATTC	TCTTATTACT	GGCTCGAAAA	TGCCTCTGCC	TAAATTACAT 3720
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT 3780
3781	ACTGGTAAGA	ATTTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT 3840

FIG. 6-1

8 / 16

3841	TCCGGTGTTT	ATTCTTATTT	AACGCCTTAT	TTATCACACG	GTCGGTATTT	CAAACCATTA	3900
3901	AATTTAGGTC	AGAAGATGAA	ATTAACATAA	ATATATTTGA	AAAAGTTTTT	TCGCGTTCTT	3960
3961	TGCTTTGCGA	TTGGATTTGC	ATCAGCATTT	ACATATAGTT	ATATAACCCA	ACCTAAGCCG	4020
4021	GAGGTTAAAA	AGGTAGTCTC	TCAGACCTAT	GATTTTGATA	AATTCACAT	TGACTCTTCT	4080
4081	CAGCGCTTTA	ATCTAAGCTA	TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAAT	4140
4141	AGCGACGATT	TACAGAAGCA	AGGTATTATCA	CTCATATATA	TTGATTTATG	TACTGTTTCC	4200
4201	ATTAATAAAG	GTAATTCAAA	TGAAATTTGTT	AAATGTAATT	AAATTTGTTT	TCTTGATGTT	4260
4261	TGTTTCATCA	TCCTCTTTTG	CTCAGGTAAT	TGAAATGAAT	AATTCGCCTC	TGCGCGATT	4320
4321	TGTAACCTGG	TATTCAAAGC	AATCAGGCGA	ATCCGTTATT	GTTTCTCCCG	ATGTAAAAGG	4380
4381	TACTGTACT	GTATATTCAT	CTGACGTTAA	ACCTGAAAAA	CTACGCAATT	TCTTTATTTT	4440
4441	TGTTTACGT	GCTAATAATT	TTGATATGGT	TGGTTCAATT	CCTTCCATAA	TTCAGAAGTA	4500
4501	TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560
4561	TGATAATTCC	GCTCCTTCTG	GTGGTTTCTT	TGTTCCGCAA	AATGATAATG	TTACTCAAAC	4620
4621	TTTTAAAAAT	AATAACGTTT	GGGCAAAAGGA	TTTAATACGA	GTTGTGCAAT	TGTTTGTAAT	4680
4681	GTCTAATACT	TCTAAATCCT	CAAAATGATT	ATCTATTGAC	GGCTCTAATC	TATTAGTTGT	4740
4741	TAGTGCACCT	AAAGATATTT	TAGATAACCT	TCCTCAATTC	CTTTCTACTG	TTGATTGGCC	4800
4801	AACTGACCAG	ATATTGATTG	AGGGTTTGAT	ATTTGAGGTT	CAGCAAGGTG	ATGCTTTAGA	4860
4861	TTTTTCATTT	GCTGCTGGCT	CTCAGCGTGG	CAGTGTGGCA	GGCGGTGTTA	ATACTGACCG	4920
4921	CCTCACCTCT	GTCTGCTGGT	CTGCTGGTGG	TTCTGTTGCA	ATTTTAAATG	GCGATGTTTT	4980
4981	AGGGCTATCA	GTTCGCGCAT	TAAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	5040
5041	TATTTCTACG	CTTTCAGGTC	AGAAGGGTTC	TATCTCTGTT	GGCCAGAATG	TCCCTTTTAT	5100
5101	TACTGGTCGT	GTGACTGGTG	AATCTGCCAA	TGTAATAAAT	CCATTTTACA	CGATTGAGCG	5160
5161	TCAAAATGTA	GGTATTTCCA	TGAGCGTTTT	TCCTGTTGCA	ATGGCTGGCG	GTAATATTGT	5220
5221	TCTGGATATT	ACCAGCAAGG	CCGATAGTTT	GAGTCTTCTT	ACTCAGGCAA	GTGATGTTAT	5280
5281	TACTAATCAA	AGAAGTATTG	CTACAACGGT	TAATTTGCGT	GATGGACAGA	CTCTTTTACT	5340
5341	CGGTGGCCTC	ACTGATTATA	AAAACACTTC	TCAAGATTCT	GGCGTACCGT	TCCGTGCTAA	5400
5401	AATCCCTTTA	ATCGGCCTCC	TGTTTAGCTC	CCGCTCTGAT	TCCAACGAGG	AAAGCACGTT	5460
5461	ATACGTGCTC	GTCAAAGCAA	CCATAGTAGC	CGCCCTGTAG	CGGCGCATT	AGCGCGGCGG	5520
5521	GTGTGGTGGT	TACGCGCAGC	GTGACCGCTA	CACTTGCCAG	CGCCCTAGCG	CCCCTCCTT	5580
5581	TCGCTTTCTT	CCCTTCCTTT	CTCGCCACGT	TCGCCCCTTT	TCCCCGTCAA	GCTCTAAATC	5640
5641	GGGGGCTCCC	TTTAGGGTTC	CGATTTAGTG	CTTTACGGCA	CCTCGACCCC	AAAAAACTTG	5700
5701	ATTTGGGTGA	TGGTTACAGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTTT	CGCCCTTTGA	5760
5761	CGTTGGAGTC	CACGTTCTTT	AATAGTGGAC	TCTTGTTCCA	AACTGGAACA	ACACTCAACC	5820
5821	CTATCTCGGG	CTATTCTTTT	GATTTATAAG	GGATTTTGCC	GATTTTCGAA	CCACCATCAA	5880
5881	ACAGGATTTT	CGCCTGCTGG	GGCAAACCG	CGTGGACCGC	TTGCTGCAAC	TCTCTCAGGG	5940
5941	CGAGGCGGTG	AAGGGCAATC	AGCTGTTGCC	CGTCTCGCTG	GTGAAAAGAA	AAACCACCTT	6000
6001	GGCGCCCAAT	ACGCAACCGG	CCTCTCCCCG	CGCGTTGGCC	GATTCATTAA	TGCAGCTGGC	6060
6061	ACGACAGGTT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	CGCAATTAAT	GTGAGTTAGC	6120
6121	TCACTCATTA	GGCACCCTAG	GCTTTACACT	TTATGCTTCC	GGCTCGTATG	TTGTGTGGAA	6180
6181	TTGTGAGCGG	ATAACAATTT	CACACGCCAA	GGAGACAGTC	ATAATGAAAT	ACCTATTGGC	6240
6241	TACGGCAGCC	GCTGGATTGT	TATTACTCGC	TGCCCAACCA	GCCATGGCCG	AGCTCGTGAT	6300
6301	GACCCAGACT	CCAGAATTCC	ATCCGGAATG	AGTGTAAATT	CTAGAACGCG	TAAGCTTGGC	6360
6361	ACTGGCCGTC	GTTTTACAAC	GTCGTGACTG	GGAAAACCTT	GGCGTTACCC	AACTTAATCG	6420
6421	CCTTGACGCA	CACCCCTTCT	TCGCCAGCTG	GCCTGAATGG	GAAGAGGCCC	GCACCGATCG	6480
6481	CCCTTCCCAA	CAGTTGCGCA	GGCTGGAGTG	CGATTTTCTT	TTTGCCTGGT	TTCCGGCACC	6540
6541	AGAAGCGGTG	CCGGAAAGCT	ACGGTTACGA	TGCGCCCATC	GAGGCCGATA	CGGTCTGCTG	6600
6601	CCCCTCAAAC	TGGCAGATGC	TTGTTCCAC	GGAGAATCCG	TACACCAACG	TAACCTATCC	6660
6661	CATTACGGTC	AATCCGCCGT	GGCTACAGGA	AGGCCAGACG	TCCGGTTGTT	ACTCGCTCAC	6720
6721	ATTTAATGTT	GATGAAAGCT	GCTGATTTAA	CAAAAATTTA	CGAATTATTT	TTGATGGCGT	6780
6781	TCCTATTGGT	TAAAAAATGA	ATTTGCTTAT	ACAATCTTCC	ACGCGAATTT	TAACAAAATA	6840
6841	TTAACGTTTA	CAATTTAAAT	GATTGACATG	CTAGTTTTAC	TGTTTTTGGG	GCTTTTCTGA	6900
6901	TTATCAACCG	GGGTACATAT	AGGCAATGAC	CTGATAGCCT	GATTACCGTT	CATCGATTCT	6960
6961	CTTGTTGCT	CCAGACTCTC	TTTATCAGCT	AGAACGGTTG	TTGTAGATCT	CTCAAAAATA	7020
7021	GTTACCTCT	CCGGCATTAA	TCACCCCTTT	GAATCTTTAC	AATATCATAT	TGATGGTGAT	7080
7081	TTGACTGTCT	CCGGCCTTTT	TTCTAAAAAT	TTTATCCTTT	CTACACATTA	CTCAGGCATT	7140
7141	GCAATTAATA	TATATGAGGG	TCATAATGTT	TTTGGTACAA	GCGTTGAAAT	AAAGGCTTCT	7200
7201	CCCGCAAAAG	TATTACAGGG	TGCTAATTTT	TTGCTTGCC	CCGATTTAGC	TTTATGCTCT	7260
7261	GAGGCTTTAT	TGCTAATTTT	TGCTAATTTT	TTGCTTGCC	TGTATGATTT	ATTGGACGTT	7320

FIG. 6-2

9 / 16

	1	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAAT	60
61	ATAGCTAAAC	AGGTTATTGA	CCATTTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT	120
121	CGTTGCGAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA	180
181	GTTGCAATAT	TAAAAACATG	TGAGCTACAG	CACCAGATTC	AGCAATTAAG	CTCTAAGCCA	240
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAGG	TACTCTCTAA	TCTTGACCTG	300
301	TTGGAGTTTG	CTTCCGGTCT	GGTTTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTTGAAG	360
361	TCTTTCGGGC	TTCTCTTTAA	TCTTTTTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT	420
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGAACT	GTITAAAGCA	480
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT	540
541	AAACATTTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTTG	CAAAAGCCTC	TCGCTATTTT	600
601	GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCCTCGT	660
661	AATTCCTTTT	GGCGTTATGT	ATCTGCATTA	GTGAAATGTG	GTATTCCTAA	ATCTGACCTG	720
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATTTT	780
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA	840
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGTTC	900
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG	960
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC	1020
1021	TGTACACCGT	TCATCTGTCC	TCTTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC	1080
1081	GCTGCGCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTCTG	CGGATTTCTGA	CACAATTTAT	1140
1141	CAGGCGATGA	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT	1200
1201	CAAAGATGAG	TGTTTTAGTG	TATTTCTTCG	CCTCTTTTCG	TTTAGGTTGG	TGCCTTCGTA	1260
1261	GTGGCATTAC	GTATTTTACC	CGTTTAATGG	AAACTTCCTC	ATGAAAAAGT	CTTTAGTCTT	1320
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCTCGT	TCCGATGCTG	TCTTTCGCTG	CTGAGGGTGA	1380
1381	CGATCCCGCA	AAAGCGGCCT	TTAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440
1441	TGCGTGGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA	1500
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAT	GGCTCCTTTT	GGAGCCTTTT	1560
1561	TTTTTGAGAG	TTTTCAACGT	GAAAAAATTA	TTATTCGCAA	TTCTTTTAGT	TGTTCTTTTC	1620
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAAATTA	1680
1681	TTTACTAAGC	TCTGGAAGAA	CGACAAAACT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
1741	CTGTGGAATG	CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA	1800
1801	TGGGTTCCCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920
1921	ATTCGGGGCT	ATACTTATAT	CAACCCTCTC	GACGGCACTT	ATCCGCCTGG	TACTGAGCAA	1980
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	2040
2041	CAGAATAATA	GGTTCGGAAA	TAGGCAGGGG	GCATTAAGTG	TTTATACGGG	CAGTGTACTT	2100
2101	CAAGGCATCG	ACCCCGTTAA	AACCTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG	2160
2161	TATGACGCTT	ACTGGAACGG	TAAATTCAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA	2220
2221	GATCCATTCTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT	2280
2281	GCTGGCGGGC	GCTCTGGTGG	TGGTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTCCGGT	2400
2401	GATTTTGATT	ATGAAAAGAT	GGCAACCGCT	ATAAGGGGGG	CTATGACCGA	AAATCCGGAT	2460
2461	GAAGAACGCG	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTGCTAC	TGATTACGGT	2520
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	2580
2581	GGTGATTTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCACCT	2640
2641	TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT	2700
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTTTCTATTG	ATTGTGACAA	AATAAACTTA	2760
2761	TTCCGTGGTG	TCTTTGCGTT	TCTTTTATAT	TATGCCACCT	TTATGTATGT	ATTTTCTACG	2820
2821	TTTGCTAACA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTCTTTTGG	GGTATTCCGT	2880
2881	TATTATTGCG	TTTCTCTGGT	TTCTTCTGG	TAACTTTGTT	CGGCTATCTG	CTTACTTTTC	2940
2941	TTAAAAAGGG	CTTCCGTAAG	ATAGCTATTG	CTATTTCAAT	GTITCTTGCT	CTTATTATTG	3000
3001	GGCTTAACTC	AATTCTTGTT	GGTTATCTCT	CTGATATTAG	CGCTCAATTA	CCCTCTGACT	3060
3061	TTGTTCAAGG	TGTTTCAGTTA	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTTT	TATGTTATTTC	3120
3121	TCTCTGTAAA	GGCTGCTATT	TTCAATTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTTGG	3180
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAAGTGGCA	AATTAGGCTC	TGGAAAGACG	3240
3241	CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT	3300
3301	CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAAC	GCCTCGCGTT	3360
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGGCG	CGGTAATGAT	3420
3421	TCCTACGATG	AAAATAAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGCGGTAC	TTGGTTTAA	3480
3481	ACCCGTTCTT	GGAAATGATA	GCGATTATTG	CTTATTATTG	ATTGGTTTCT	ACATGCTCGT	3540
3541	AAATTAGGAT	GGGATATTAT	TTTTCTTGTT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG	3600
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTTAT	TGTCGTCGTC	TGGACAGAAT	TACTTTACCT	3660
3661	TTTGTGCGTA	CTTTATATTC	TCTTATTACT	GGCTCGAAAA	TGCCTCTGCC	TAAATTACAT	3720
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT	3780

FIG. 7-1

10 / 16

3781	ACTGGTAAGA	ATTTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT	3840
3841	TCCGGTGTTT	ATTCTTATTT	AACGCCTTAT	TTATCACACG	GTCGGTATTT	CAAACCATTA	3900
3901	AATTTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATATTTGA	AAAAGTTTTT	ACGCGTTCTT	3960
3961	TGTCTTGCGA	TTGGATTGTC	ATCAGCATTT	ACATATAGTT	ATATAACCCA	ACCTAAGCCG	4020
4021	GAGGTTAAAA	AGGTAGTCTC	TCAGACCTAT	GATTTTGATA	AATTCACTAT	TGACTCTTCT	4080
4081	CAGCGTCTTA	ATCTAAGCTA	TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAAT	4140
4141	AGCGACGATT	TACAGAAGCA	AGGTTATTCA	CTCACATATA	TTGATTTATG	TACTGTTTCC	4200
4201	ATTAATAAAG	GTAATTCAAA	TGAAATGTTT	AATGTAAATT	AATTTTGTTT	TCTTGATGTT	4260
4261	TGTTTCATCA	TCTTCTTTTG	CTCAGGTAAT	TGAAATGAAT	AATTCGCCTC	TGCGCGATTG	4320
4321	TGTAACCTGG	TATTCAAAGC	AATCAGGCGA	ATCCGTTATT	GTTTCTCCCG	ATGTAATAAG	4380
4381	TACTGTTACT	GTATATTCAT	CTGACGTTAA	ACCTGAAAT	CTACGCAATT	TCTTTATTTT	4440
4441	TGTTTTACGT	GCTAATAAAT	TTGATATGGT	TGGTTCAATT	CCTTCCATAA	TTCAGAAGTA	4500
4501	TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560
4561	TGATAATTC	GCTCCTTCTG	GTGGTTTCTT	TGTTCCGCAA	AATGATAATG	TTACTCAAAC	4620
4621	TTTTAAAAAT	AATAACGTTT	GGGCAAGGGA	TTTAATACGA	GTTGTGCAAT	TGTTTGTAAG	4680
4681	GTCTAATACT	TCTAAATCCT	CAAAATGTAT	ATCTATTGAC	GGCTCTAATC	TATTAGTTGT	4740
4741	TAGTGCACCT	AAAGATATTT	TAGATAACCT	TCCTCAATTC	CTTTCTACTG	TTGATTTGCC	4800
4801	AACTGACCAG	ATATTGATTG	AGGGTTTGAT	ATTGAGGTT	CAGCAAGGTG	ATGCTTTAGA	4860
4861	TTTTTCATTT	GCTGCTGGCT	CTCAGCGTGG	CACTGTTGCA	GGCGGTGTTA	ATACTGACCG	4920
4921	CCTCACCTCT	GTTTTATCTT	CTGCTGGTGG	TTGTTCCGGT	ATTTTTAATG	GCGATGTTTT	4980
4981	AGGGCTATCA	GTTTCGCGCAT	TAAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	5040
5041	TATTCCTACG	CTTTTCAGGTG	AGAAGGGTTC	TATCTCTGTT	GGCCAGAATG	TCCCTTTTAT	5100
5101	TACTGGTCGT	GTGACTGGTG	AATCTGCCAA	TGTAATAAAT	CCATTTTACA	CGATTGACGG	5160
5161	TCAAAATGTA	GGTATTTCCA	TGAGCGTTTT	TCCTGTTGCA	ATGGCTGGCG	GTAATATTGT	5220
5221	TCTGGATATT	ACCAGCAAGG	CCGATAGTTT	GAGTCTTCT	ACTCAGGCAA	GTGATGTTAT	5280
5281	TACTAATCAA	AGAAGTATTG	CTACAACGGT	TAATTTGCGT	GATGGACAGA	CTCTTTTACT	5340
5341	CGGTGGCCTC	ACTGATTATA	AAAACACTTC	TCAAGATTCT	GGCGTACCGT	TCCTGTCTAA	5400
5401	AATCCCTTTA	ATCGGCCTCC	TGTTTAGTCT	CCGCTCTGAT	TCCAACGAGG	AAAGCACGTT	5460
5461	ATACGTGCTC	GTCAAAGCAA	CCATAGTACG	CGCCCTGTAG	CGGCGCATT	AGCGCGCGGG	5520
5521	GTGTGGTGGT	TACGCGCAGC	GTGACCGCTA	CACTTGCCAG	CGCCCTAGCG	CCCGCTCCTT	5580
5581	TGCTTTTCTT	CCCTTCCTTT	CTCGCCACGT	TGCGCCGGCT	TCCCCGTCAA	GCTCTAAATC	5640
5641	GGGGGCTCCC	TTTAGGGTTC	CGATTTAGTG	CTTTACGGCA	CCTCGACCCC	AAAAAACTTG	5700
5701	ATTTGGGTGA	TGGTTCACGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTTT	CGCCCTTTGA	5760
5761	CGTTGGAGTC	CACGTTCTTT	AATAGTGGAC	TCTTGTTCCA	AACTGGAACA	ACACTCAACC	5820
5821	CTATCTCGGG	CTATTTCTTT	GATTATAAAG	GGATTTTGCC	GATTTGCGAA	CCACCATCAA	5880
5881	ACAGGATTTT	CGCCTGTGTT	GGCAAACCAG	CGTGGACCGC	TTGCTGCAAC	TCTCTCAGGG	5940
5941	CCAGGCGGTT	AAGGGCAATC	AGCTGTTGCC	CGTCTCGCTG	GTGAAAAGAA	AAACCACCCT	6000
6001	GGCGCCCAAT	ACGCAAACCG	CCTCTCCCGG	CGCGTTGGCC	GATTCATTAA	TGCAGCTGGC	6060
6061	ACGACAGGTT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	CGCAATTAAT	GTGAGTTAGC	6120
6121	TCACTCATTA	GGCACCCTAG	GCTTTACACT	TTATGCTTCC	GGCTCGTATG	TTGTGTGGAA	6180
6181	TTGTGAGCGG	ATAACAATTT	CACACGCGTC	ACTTGGCACT	GGCCGTGCTT	TTACAACGTC	6240
6241	GTGACTGGGA	AAACCTTGGC	GTTACCCAAG	CTTTGTACAT	GGAGAAAATA	AAGTGAAACA	6300
6301	AAGCACTATT	GCACTGGCAC	TCTTACCGTT	ACCGTTACTG	TTTACCCCTG	TGACAAAAGC	6360
6361	CGCCCAAGTC	CAGCTGCTCG	AGTCAGGCTT	ATTGTGCCCA	GGGGATTGTA	CTAGTGGATC	6420
6421	CTAGGCTGAA	GGCGATGACC	CTGCTAAGGC	TGCATTCAAT	AGTTTACAGG	CAAGTGCTAC	6480
6481	TGAGTACATT	GGCTACGCTT	GGGCTATGGT	AGTAGTTATA	GTTGGTGCTA	CCATAGGGAT	6540
6541	TAAATTATTC	AAAAAGTTTA	CGAGCAAGGC	TTCTTAAGCA	ATAGCGAAGA	GGCCCGCACC	6600
6601	GATCGCCCTT	CCCAACAGTT	GCAGCAGCTG	AATGGCGAAT	GGCGCTTTGC	CTGGTTTCCG	6660
6661	GCACCAGAAG	CGGTGCCGGA	AAGCTGGCTG	GAGTGCGATC	TTCTTGAGGC	CGATACGGTC	6720
6721	GTCTGCCCTT	CAAACTGGCA	GATGCACGGT	TACGATGCGC	CCATCTACAC	CAACGTAACC	6780
6781	TATCCCATT	CGGTCAATCC	GCCGTTTGT	CCCACGGAGA	ATCCGACGGG	TTGTTACTCG	6840
6841	CTCACATTTA	ATGTTGATGA	AAGCTGGCTA	CAGGAAGGCC	AGACGCGAAT	TATTTTGTAT	6900
6901	GGCGTTCCTA	TTGGTTAAAA	AATGAGCTGA	TTTAACAAAA	ATTTAACGCG	AAATTTAACA	6960
6961	AAATATTAAC	GTTTACAATT	TAAATATTTG	CTTATACAAT	CTTCTGTTT	TTGGGGCTTT	7020
7021	TCTGATTATC	AACCGGGGTA	CATATGATTG	ACATGCTAGT	TTTACGATTA	CCGTTTCATC	7080
7081	ATTCTCTTGT	TTGCTCCAGA	CTCTCAGGCA	ATGACCTGAT	AGCCTTTGTA	GATCTCTCAA	7140
7141	AAATAGCTAC	CCTCTCCGGC	ATTAATTTAT	CAGCTAGAAC	GGTTGAATAT	CATATTGATG	7200
7201	GTGATTGAC	TGTCTCCGGC	CTTTCTCACC	CTTTTGAATC	TTTACCTACA	CATTACTCAG	7260
7261	GCATTGCATT	TAAAATATAT	GAGGGTTCTA	AAAAATTTTA	TCCTTGCGTT	GAAATAAAGG	7320
7321	CTTCTCCCGC	AAAAGTATTA	CAGGGTCATA	ATGTTTTTGG	TACAACCGAT	TTAGCTTTAT	7380
7381	GCTCTGAGGC	TTTATTGCTT	AATTTTGCTA	ATTCTTTGCC	TTGCTGTAT	GATTATTGGG	7440
7441	ACGTT						7445

FIG. 7-2

SUBSTITUTE SHEET

11 / 16

	10	20	30	40	50	60	
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAAT	60
61	ATAGCTAAAC	AGGTTATTGA	CCATTTTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT	120
121	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA	180
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAGATTG	AGCAATTAAG	CTCTAAGCCA	240
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAAGGAG	CAATTAAGG	TACTCTCTAA	TCCTGACCTG	300
301	TTGGAGTTTG	CTTCCGGTCT	GGTTCGCTTT	GAAGCTCGAA	TAAAAACGCG	ATATTTGAAG	360
361	TCTTTCGGGC	TTCTCTTAA	TCTTTTTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT	420
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCAITCTCGT	TTTCTGAACT	GTTTAAAGCA	480
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT	540
541	AAACATTTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTTG	CAAAAAGCCTC	TCGCTATTTT	600
601	GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCTCGT	660
661	AATTCCTTTT	GGCGTTATGT	ATCTGCATTA	TGTGAATGTG	GTATTCCTAA	ATCTCAACTG	720
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATTTT	780
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCCTA	AAATCGCATA	AGGTAATTCA	840
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGTGT	900
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG	960
961	AATATCCGGT	TCTTGCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	CGCCTAGGCT	1020
1021	TGTACACCGT	TCTATCTGTC	TCTTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC	1080
1081	GTCTGCGCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTCT	CGGATTTCTG	CACAATTTAT	1140
1141	CAGGCGATGA	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT	1200
1201	CAAAGATGAG	TGTTTTAGTG	TATTTCTTTC	CCTCTTTCGT	TTTAGGTTGG	TGCCTTCGTA	1260
1261	GTGGCATTAC	GTATTTTACC	CGTTTAAATG	AAACTTCCTC	ATGAAAAAGT	CTTTAGTCTT	1320
1321	CAAAAGCCTCT	GTAGCCGTTG	CTACCCTCGT	TCCGATGCTG	TCTTTTCGCTG	CTGAGGGTGA	1380
1381	CGATCCCGCA	AAAGCGGCCT	TAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440
1441	TGCGTGGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAAGAA	1500
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAG	GGCTCCTTTT	GGAGCCTTTT	1560
1561	TTTTTGAGAG	TTTTCAACGT	GAAAAAATTA	TTATTCGCAA	TTCTTTTAGT	TGTTCCCTTC	1620
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AAACCCATAC	AGAAAAATCA	1680
1681	TTTACTAACG	CTTGAAAGA	CGACAAAAC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
1741	CTGTGGAATG	CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA	1800
1801	TGGGTTTCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920
1921	ATTCGGGGCT	ATACTTATAT	CAACCTCTC	GACGGCACTT	ATCCGCCTGG	TACTGAGCAA	1980
1981	AAACCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCTGTTT	2040
2041	CAGAATAATA	AGTTCCGAAA	TAGGCAAGGG	GCAATTAACG	TTTATACGGG	CAGTGTACT	2100
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG	2160
2161	TATGACGCTT	ACTGGAACGG	TAAATTCAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA	2220
2221	GATCCATTCT	TTTGTGAATA	TCAAGGCCAA	TGCTCTGACC	TGCCTCAACC	TCCTGTCAAT	2280
2281	GCTGGCGGCG	CGTCTGGTGG	TGGTCTGTTG	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGAT	2340
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTTCGGGT	2400
2401	GATTTTGATT	ATGAAAAGAT	GGCAAAAGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460
2461	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTGCTAC	TGATTACGGT	2520
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	2580
2581	GGTGATTTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCACCT	2640
2641	TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT	2700
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATTGA	TTTCTATTG	ATTGTGACAA	AGTAAACTTA	2760
2761	TTCCGTGGTG	TCTTTGCGTT	TCTTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTTCTACG	2820
2821	TTTGCTAACA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTTG	GGTATTCCGT	2880
2881	TATTATTGCG	TTTCCTCGGT	TTCTTCTGG	TAACCTTGT	C6GCTATCTG	CTTACTTTTC	2940
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTTATT	GTTTCTTGCT	CTTATTATTG	3000
3001	GGCTTAACTC	AATTCCTTGT	GGTTATCTCT	CTGATATTAG	CGCTCAATTA	CCCTCTGACT	3060
3061	TTGTTACAGG	TGTTACGTTA	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTTT	TATGTTATTC	3120
3121	TCTCTGTAAA	GGCTGCTATT	TTCATTTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTGG	3180
3181	ATTGGGATAA	ATAATATGGC	TGTTTTATT	GTAAGTGGCA	AATTAGGCTC	TGGAAGACG	3240
3241	CTCGTTAGCG	TTGGTAAGAT	TTAGGATAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT	3300
3301	CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAC	GCCTCGCGTT	3360
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTCGTTG	CTATTGGGCG	CGGTAATGAT	3420
3421	TCCTACGATG	AAAAATAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGC6GTAC	TTGGTTTAA	3480
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTTCT	ACATGCTCGT	3540
3541	AAATTAGGAT	GGGATATTAT	TTTTCTTGTT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG	3600
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTTAT	TGTCGTCGTC	TGGACAGAAT	TACTTTACCT	3660
3661	TTTGTGCGTA	CTTTATATTC	TCTTATTACT	GGCTCGAAAA	TGCCTCTGCC	TAAATTACAT	3720
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT	3780
3781	ACTGGTAAGA	ATTTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT	3840

FIG. 8-1

SUBSTITUTE SHEET

12 / 16

3841	TCCGGTGT	ATTCTTATT	AACGCCTTAT	TTATCACACG	GTCGGTATTT	CAAACCATTA	3900
3901	AATTTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATATTTGA	AAAAGTTTTT	ACGCGTTCTT	3960
3961	TGCTTTGCGA	TTGGATTTGC	ATCAGCATTT	ACATATAGTT	ATATAACCCA	ACCTAAGCCG	4020
4021	GAGGTAAAA	AGGTAGTCTC	TCAGACCTAT	GATTTTGATA	AATTCATAT	TGACTCTTCT	4080
4081	CAGCGTCTTA	ATCTAAGCTA	TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAAT	4140
4141	AGCGACGATT	TACAGAAGCA	AGGTTATTCA	CTCACATATA	TTGATTTATG	TACTGTTTCC	4200
4201	ATTAAAAAAG	GTAATTCAAA	TGAAATTTGT	AAATGTAATT	AATTTTGTTT	TCTTGATGTT	4260
4261	TGTTTCATCA	TCTTCTTTTG	CTCAGGTAAT	TGAAATGAAT	AATTCGCCTC	TGCGCGATTT	4320
4321	TGTAACCTGG	TATTCAAAGC	AATCAGGCGA	ATCCGTTATT	GTTTCTCCCG	ATGTAAGAGG	4380
4381	TACTGTTACT	GTATATTCAT	CTGACGTTAA	ACCTGAAAAT	CTACGCAATT	TCTTTATTTT	4440
4441	TGTTTTACGT	GCTAATAATT	TTGATATGGT	TGGTTCAATT	CCTTCCATAA	TTCAGAAGTA	4500
4501	TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560
4561	TGATAATTCC	GCTCCTTCTG	GTGGTTTCTT	TGTTCCGCAA	AATGATAATG	TTACTCAAAG	4620
4621	TTTTAAAAAT	AATAACGTTT	GGGCAAAAGG	TTTAATACGA	GGCTCTAATC	TGTTTGTAAG	4680
4681	GTCTAATACT	TCTAAATCCT	CAAAATGTAT	ATCTATTGAC	CTTCTACTG	TATTAGTTGT	4740
4741	TAGTGCACCT	AAAGATATTT	TAGATAACCT	TCCTCAATTC	CAGCAAGGTG	TTGATTTGCC	4800
4801	AACTGACCAG	ATATTGATTG	AGGGTTTGAT	ATTTGAGGTT	GGCGGTGTTA	ATGCTTTAGA	4860
4861	TTTTTCATTT	GCTGCTGGCT	CTCAGCGTGG	CAGTGTGCGA	ATTTTTAATG	ATACTGACCG	4920
4921	CCTCACCTCT	TTTTATCTTT	CTGCTGGTGG	TTTCTGCGT	AAAATATTGT	GCGATGTTTT	4980
4981	AGGGCTATCA	GTTCGCGCAT	TAAAGACTAA	TAGCCATTCA	GGCCAGAATG	CTGTGCCACG	5040
5041	TATTTCTACG	CTTTCAGGTC	AGAAGGGTTC	TATCTCTGTT	CCATTTTACA	TCCCTTTTAT	5100
5101	TACTGGTCGT	GTGACTGGTG	AATCTGCCAA	TGTAATAAAT	ATGGCTGGCG	CGATTGAGCG	5160
5161	TCAAAATGTA	GGTATTTCCA	TGAGCGTTTT	TCCTGTTGCA	ACTCAGGCAA	GTAATATTGT	5220
5221	TCTGGATATT	ACCAGCAAGG	CCGATAGTTT	GAGTCTTCTT	GATGGACAGA	GTGATGTTAT	5280
5281	TACTAATCAA	AGAAGTATTG	CTACAACGGT	TAATTTGCGT	GGCGTACCGT	CTCTTTTACT	5340
5341	CGGTGGCCTC	ACTGATTATA	AAAACACTTC	TCAAGATTCT	TCCAACGAGG	TCCCTGTCTAA	5400
5401	AATCCCTTTA	ATCGGCCTCC	TGTTTAGCTC	CCGCTCTGAT	CGGCGCATTA	AAAGCACGTT	5460
5461	ATACGTGCTC	GTCAAAGCAA	CCATAGTACG	CGCCCTGTAG	CGCCCTAGCG	AGCGCGGCGG	5520
5521	GTGTGGTGGT	TACGCGCAGC	GTGACCGCTA	CACCTGGCCAG	TCCCGTCAA	CCCGCTCCTT	5580
5581	TCGCTTTCTT	CCCTTCCTTT	CTCGCCACGT	TCGCGGCTT	CCTCGACCCC	GCTCTAAATC	5640
5641	GGGGGCTCCC	TTTAGGGTTC	CGATTTAGTG	CTTTACGGCA	GACGGTTTTT	AAAAAACTTG	5700
5701	ATTTGGGTGA	TGGTTCACGT	AGTGGGCCAT	CGCCCTGATA	AACTGGAACA	CGCCCTTTGA	5760
5761	CGTTGGAGTC	CACGTTCTTT	AATAGTGGAC	TCTTGTTCCA	GATTTCGGAA	ACACTCAACC	5820
5821	CTATCTCGGG	CTATTCTTTT	GATTTATAAG	GGATTTGCC	TTGCTGCAAC	CCACCATCAA	5880
5881	ACAGGATTTT	CGCCTGCTGG	GGCAAACCGA	CGTGGACCGC	GTGAAAAGAA	TCTCTCAGGG	5940
5941	CCAGGCGGTG	AAGGGCAATC	AGCTGTTGCC	CGTCTCGCTG	GATTCAATTA	AAACCACCTG	6000
6001	GGCGCCCAAT	ACGCAAAACG	CCTCTCCCGG	CGCGTTGGCC	CGCAATTAAT	TGCAGCTGGC	6060
6061	ACGACAGGTT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	GGCTCGTATG	GTGAGTTAGC	6120
6121	TCACCTATTA	GGCACCCAG	GCTTTACACT	TTATGCTTCC	GGCCGTCGTT	TTGTTGGGAA	6180
6181	TTGTGAGCGG	ATAACAATTT	CACACGCGTC	ACTTGGCACT	GGAGAAAATA	TTACAACGTC	6240
6241	GTGACTGGGA	AAACCCTGGC	GTTACCCAAG	CTTTGTACAT	CCTGTGGCAA	AAGTGAACAA	6300
6301	AAGCACTATT	GCACCTGGAC	TCTTACCGTT	ACTGTTTACC	GACCCTGCTA	AAGCCTATGG	6360
6361	GGGGTTTCATG	CTTCTGAGGC	ATCCGGGAGC	TGAAGGCGAT	GCTTGGGCTA	AGGCTGCATT	6420
6421	CAATAGTTTA	CAGGCAAGTG	CTACTGAGTA	CATTGGCTAC	TTTACGAGCA	TGGTAGTAGT	6480
6481	TATAGTTGGT	GCTACCATAG	GGATTAATTA	ATTCAAAAAG	AGTTGCGCAG	AGGCTTCTTA	6540
6541	AGCAATAGCG	AAGAGGCCCC	CACCGATCGC	CCTTCCCAAC	CGGAAAGCTG	CCTGAATGGC	6600
6601	GAATGGCGCT	TTGCCTGGTT	TCCGGCACCA	GAAGCGGTGC	GCTGGAGTGC	GCTGGAGTGC	6660
6661	GATCTTCCTG	AGGCCGATAC	GGTCGTCGTC	CCCTCAAAC	GGCAGATGCA	CGGTTACGAT	6720
6721	GCGCCCATCT	ACACCAACGT	AACCTATCCC	ATTACGGTCA	ATCCGCCGTT	TGTTCCACAG	6780
6781	GAGAATCCGA	CGGGTTGTTA	TGATGGCGTT	TTTAATGTTG	ATGAAAAGCTG	GCTACAGGAA	6840
6841	GGCCAGACGG	GAATTTATTT	AACAAAATAT	CTTATGTTG	AAAAAATGAG	CTGATTTAAC	6900
6901	AAAAATTTAA	CGCGAATTTT	CTTTTCTGAT	TAACGTTTAC	AATTTAAATA	TTTGCTTATA	6960
6961	CAATCTTCTT	TTTTTTGGGG	ATCGATTCTC	TATCAACCGG	GGTACATATG	ATTGACATGC	7020
7021	TAGTTTACG	ATTACCGTTC	TCAAAAATAG	TTGTTTGCTC	CAGACTCTCA	GGCAATGACC	7080
7081	TGATAGCCTT	TGTAGATCTC	GATGGTGATT	CTACCCTCTC	CGGCATTAA	TTATCAGCTA	7140
7141	GAACGGTTGA	ATATCATATT	TCAGGCATTG	TGACTGTCTC	CGGCCTTTCT	CACCCCTTTG	7200
7201	AATCTTTACC	CGTTGAAATA	AAGGCTTCTC	CATTTAAAT	ATATGAGGGT	TCTAAAAAT	7260
7261	TTTATCTTGG	CGATTAGCT	TTATGCTCTG	AGGCTTTATT	ATTACAGGGT	CATAATGTTT	7320
7321	TTGGTACAAC	GTATGATTTA	TTGGACGTT		GCTTAATTTT	GCTAATTTCT	7380
7381	TGCCCTGCCT						7409

FIG. 8-2

13 / 16

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAAT 60
61	ATAGCTAAAC	AGGTTATTGA	CCATTTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT 120
121	CGTTGCGAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA 180
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAGATTG	AGCAATTAAG	CTCTAAGCCA 240
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAAGGAG	CAATTAAGG	TACTCTCTAA	TCCTGACCTG 300
301	TTGGAGTTTG	CTTCCGGTCT	GGTTCGCTTT	GAAGCTCGAA	TTAAAACGCG	ATATTTGAAG 360
361	TCTTTGCGGC	TTCTCTTAA	TCTTTTGTAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT 420
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA 480
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT 540
541	AAACATTTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTTG	CAAAAGCCTC	TCGCTATTTT 600
601	GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCTCGT 660
661	AATTTCTTTT	GGCGTTATGT	ATCTGCAATTA	GTGAATGTG	GTATTCCTAA	ATCTCAACTG 720
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATTTT 780
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCCTA	AAATCGCATA	AGGTAATTCA 840
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGTTT 900
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG 960
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCCAGCCTAT 1020
1021	TGTACACCGT	TCATCTGTCC	TCTTTCAAAG	TGTGTCAGTT	CGGTTCCCTT	ATGATTGACC 1080
1081	GTCTGCGCCT	CGTTCGGGCT	AAGTAACATG	GAGCAGGTCTG	CGGATTTCTA	CACAAATTTAT 1140
1141	CAGGCGATGA	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT 1200
1201	CAAAGATGAG	TGTTTTAGTG	TATTTCTTCG	CCTCTTTCGT	TTTAGGTTGG	TGCCTTCGTA 1260
1261	GTGGCATTAC	GTATTTTACC	CGTTTAAATG	AACTTCCTC	ATGAAAAAGT	CTTTAGTCTT 1320
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCTCGT	TCCGATGCTG	TCTTTTCGCTG	CTGAGGGTGA 1380
1381	CGATCCCGCA	AAAGCGGCCT	TAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA 1440
1441	TGCGTGGGCG	ATGGTTGTTG	TCATTGTCTG	CGCAACTATC	GGTATCAAGC	TGTTTAAAGAA 1500
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAG	GGCTCCTTTT	GGAGCCTTTT 1560
1561	TTTTTGGAGA	TTTCAACGTT	GAAAAAATTA	TTATTCGCAA	TTCTTTTAGT	TGTTCTTTTC 1620
1621	TATTTCTACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAAATTA 1680
1681	TTTACTAACG	TCTGGAAGAA	CGACAAAAGT	TTAGATCGTT	ACGCTAATTA	TGAGGGTTGT 1740
1741	CTGTGGAATG	CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA 1800
1801	TGGGTTCTTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGCTG	GTGGCTCTGA	GGGTGGCGGT 1860
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT 1920
1921	ATTCGGGGCT	ATACTTATAT	CAACCCTCTC	GACGGCACTT	ATCCGCCTGG	TACTGAGCAA 1980
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCTGTTT 2040
2041	CAGAATAATA	GGTTCCGAAA	TAGGCAAGGG	GACTTAACCTG	TTTATACGGG	CACCTGTTACT 2100
2101	CAAGGCACCTG	ACCCCGTTAA	AACCTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG 2160
2161	TATGACGCTT	ACTGGAACGG	TAAATTCAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA 2220
2221	GATCCATTCTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT 2280
2281	GCTGGCGGCG	GCTCTGGTGG	TGGTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT 2340
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGDTGGGCT	TGGTTCGGGT 2400
2401	GATTTTGATT	ATGAAAAGAT	GGCAAAAGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT 2460
2461	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AACTTGATT	CTGTCGCTAC	TGATTACGGT 2520
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT 2580
2581	GGTGATTTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCACCT 2640
2641	TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT 2700
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATTGA	TTTTCTATTG	ATTGTGACAA	AATAAACTTA 2760
2761	TTCCGTGGTG	TCTTTGCGTT	TCTTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTTCTACG 2820
2821	TTTGCTAACA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTTG	GGTATTCCTG 2880
2881	TATTATTGCG	TTTCTCTGGT	TTCTTTCTGG	TAACTTTGT	CGGCTATCTG	CTTACTTTTC 2940
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTT	CTATTTTCA	GTTCCTTGCT	CTTATTATTG 3000
3001	GGCTTAACTC	AATTTCTGTG	GGTTATCTCT	CTGATATTAG	CGCTCAATTA	CCCTCTGACT 3060
3061	TTGTTCAAGG	TGTTCAAGTTA	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTTT	TATGTTTATC 3120
3121	TCTCTGTAAA	GGCTGCTATT	TTCATTTTTG	ACGTAAACAA	AAAAATCGTT	TCTTATTTTG 3180
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAAGTGGCA	AATTAGGCTC	TGGAAGACG 3240
3241	CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT 3300
3301	CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAC	GCCTCGCGTT 3360
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGGCG	CGGTAATGAT 3420
3421	TCCTACGATG	AAAATAAAAA	CGGCTTGCTT	TGTTCTGATG	AGTGCGGTAC	TTGGTTTAA 3480
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTTCT	ACATGCTCGT 3540
3541	AAATTAGGAT	GGGATATTAT	CTTCCTTGT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG 3600
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTTAT	TGTCGTGCTC	TGGACAGAAT	TACTTTACCT 3660
3661	TTTGTCCGTA	CTTTATATTC	TCTTATTACT	GGCTCGAAAA	TGCCTCTGCC	TAAATTACAT 3720
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT 3780
3781	ACTGGTAAGA	ATTTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT 3840

FIG. 9-1

SUBSTITUTE SHEET

14 / 16

3841	TCCGGTGT	TTCTTATTT	AACGCCTTAT	TTATCACACG	GTCGGTATTT	CAAACCATTA	3900
3901	AATTTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATATTTGA	AAAAGTTTTG	ACGCGTTCTT	3960
3961	TGTCTTGCGA	TTGGATTTCG	ATCAGCATTT	ACATATAGTT	ATATAACCCA	ACCTAAGCCG	4020
4021	GAGGTTAAAA	AGGTAGTCTC	TCAGACCTAT	GATTTTGATA	AATTCACAT	TGACTCTTCT	4080
4081	CAGCGTCTTA	ATCTAAGCTA	TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAAT	4140
4141	AGCGACGATT	TACAGAAGCA	AGGTTATTCA	CTCACATATA	TTGATTTATG	TACTGTTTCC	4200
4201	ATTAATAAAGG	TAATTCAAAT	GAAATTGTTA	AATGTAATTA	ATTTTGTTTT	CTTGATGTTT	4260
4261	GTTTCATCAT	CTTCTTTTGC	TCAGGTAATT	GAAATGAATA	ATTCGCCTCT	GCGCGATTTT	4320
4321	GTAACCTGGT	ATTCAAAGCA	ATCAGGCGAA	TCCGTTATTG	TTTCTCCCGA	TGFAAAAGGT	4380
4381	ACTGTTACTG	TATATTCATC	TGACGTTAAA	CCTGAAAATC	TACGCAATTT	CTTTATTTCT	4440
4441	GTTTTACGTG	CTAATAATTT	TGATATGGTT	CCTTCAATTC	CTTCCATTAT	TTAGAAGTAT	4500
4501	AATCCAAACA	ATCAGGATTA	TATTGATGAA	TTGCCATCAT	CTGATAATCA	GGAATATGAT	4560
4561	GATAATTCGG	CTCCTTCTGG	TGTTTTCTTT	GTTCCGCAAA	ATGATAATGT	TACTCAAAT	4620
4621	TTTAAATAA	ATAACGTTTG	GGCAAAGGAT	TTAATACGAG	TTGTGCAATT	GTTTGTAAG	4680
4681	CTAATACTT	CTAAATCCTC	AAATGTATTA	TCTATTGACG	GCTCTAATCT	ATTAGTTGTT	4740
4741	AGTGCACCTA	AAGATATTTT	AGATAACCTT	CCTCAATTCC	TTTCTACTGT	TGATTTGCCA	4800
4801	ACTGACCAGA	TATTGATTGA	GGGTTTGATA	TTTGAGGTTT	AGCAAGGTGA	TGCTTTAGAT	4860
4861	TTTTCATTTG	CTGCTGGCTC	TCAGCTGGTG	ACTGTTGCAG	GCGGTGTTAA	TACTGACCGC	4920
4921	CTCACCTCTG	TTTTATCTTC	TGCTGGTGGT	TCGTTCCGTA	TTTTTAATGG	CGATGTTTTA	4980
4981	GGGCTATCAG	TTCCGCGCAT	AAAGACTAAT	AGCCATTCAA	AAATATTGTC	TGTGCCACGT	5040
5041	ATTCTTACGC	TTTCAGGTCA	GAAGGGTTCT	ATCTCTGTTG	GCCAGAATGT	CCCTTTTATT	5100
5101	ACTGGTCGTG	TGACTGGTGA	ATCTGCCAAT	GTAATAATC	CATTTTCAGAC	GATTGAGCGT	5160
5161	CAAAATGTAG	GTATTTCCAT	GAGCGTTTTT	CCTGTTGCAA	TGGCTGGCGG	TAATATTGTT	5220
5221	CTGGATATTA	CCAGCAAGGC	CGATAGTTTG	AGTTCTTCTA	CTCAGGCAAG	TGATGTTATT	5280
5281	ACTAATCAAA	GAAGTATTGC	TACAACGGTT	AATTTGCGTG	ATGGACAGAC	TCCTTTTACTC	5340
5341	GGTGGCCTCA	CTGATTATAA	AAACACTTCT	CAAGATTCTG	GCGTACCCTT	CCTGTCTAAA	5400
5401	ATCCCTTTAA	TCGGCCTCCT	GTTTAGCTCC	CGCTCTGATT	CCAACGAGGA	AAGCACGTTA	5460
5461	TACGTGCTCG	TCAAAGCAAC	CATAGTACGC	GCCCTGTAGC	GGCGCATTA	GCGCGGCGGG	5520
5521	TGTGGTGGTT	ACGCGCAGCG	TGACCGCTAC	ACTTGCCAGC	GCCCTAGCGC	CCGCTCCTTT	5580
5581	CGCTTTCTTC	CCTTCCTTTC	TCGCCAGCTT	GCCCGGCTTT	CCCCGTCAAG	CTCTAAATCG	5640
5641	GGGGCTCCCT	TTAGGGTTCC	GATTTAGTGC	TTTACGGCAC	CTCGACCCCA	AAAAACTTGA	5700
5701	TTTGGGTGAT	GGTTACGTA	GTGGGCCATC	GCCCTGATAG	ACGGTTTTTC	GCCCTTTGAC	5760
5761	GTTGGAGTCC	ACGTTCTTTA	ATAGTGGACT	CTTGTTCCAA	ACTGGAACAA	CACTCAACCC	5820
5821	TATCTGGGG	TATCTTTTTC	ATTTATAAGG	GATTTTGCCG	ATTTGGAAC	CACCATCAAA	5880
5881	CAGGATTTTC	GCCTGCTGGG	GCAAACCCAG	GTGGACCGCT	TGCTGCAACT	CTCTCAGGGC	5940
5941	CAGGCGGTGA	AGGGCAATCA	GCTGTTGCC	GTCTCGCTGG	TGAAAAGAAA	AACCACCCTG	6000
6001	GCGCCCAATA	CGCAAACCCG	CTCTCCCCGC	GCGTTGGCCG	ATTCATTAAT	GCAGCTGGCA	6060
6061	CGACAGGTTT	CCCGACTGGA	AAGCGGGCAG	TGAGCGCAAC	GCAATTAATG	TGAGTTAGCT	6120
6121	CACCTATTAG	GCACCCAGG	CTTTACACTT	TATGCTTCCG	GCTCGTATGT	TGTGTGGAAT	6180
6181	TGTGAGCGGA	TAACAATTTT	ACACAGGAAA	CAGCTATGAC	CAGGATGTAC	GAATTCGCAG	6240
6241	GTAGGAGAGC	TCGGCGGATC	CGAGGTGAA	GGCGATGACC	CTGCTAAGGC	TGCATTCAAT	6300
6301	AGTTTACAGG	CAAGTGCTAC	TGAGTACATT	GGCTACGCTT	GGGCTATGGT	AGTAGTTATA	6360
6361	GTTGGTGCTA	CCATAGGGAT	TAAATTATTC	AAAAAGTTTA	CGAGCAAGGC	TTCTTAACCA	6420
6421	GCTGGCGTAA	TAGCGAAGAG	GCCCCGACCG	ATCGCCCTTC	CCAACAGTTG	CGCAGCCTGA	6480
6481	ATGGCGAATG	GCGCTTTGCC	TGGTTTCCGG	CACCAGAAGC	GGTGCCGGAA	AGCTGGCTGG	6540
6541	AGTGCGATCT	TCCTGAGGCC	GATACGGTCC	TCGTCCCTTC	AAACTGGCAG	ATGCACGGTT	6600
6601	ACGATGCGCC	CATCTACACC	AACGTAACCT	ATCCCATTA	GGTCAATCCG	CCGTTTGTTC	6660
6661	CCACGGAGAA	TCCGACGGGT	TGTTACTCGC	TCACATTTAA	TGTTGATGAA	AGCTGGCTAC	6720
6721	AGGAAGGCCA	GACGCGAATT	ATTTTGTATG	GCGTTCTTAT	TGGTTAAAAA	ATGAGCTGAT	6780
6781	TTAACAATAA	TTTAACGCGA	ATTTTAAACAA	AATATTAAAG	TTTACAATTT	AAATATTTCG	6840
6841	TTATACAATC	TTCTGTTTTT	TGGGGCTTTT	CTGATTATCA	ACCGGGGTAC	ATATGATTGA	6900
6901	CATGCTAGTT	TTACGATTAC	CGTTTATCGA	TTCTCTTGT	TGCTCCAGAC	TCTCAGGCAA	6960
6961	TGACCTGATA	GCCTTTGTAG	ATCTCTCAAA	AATAGCTACC	CTCTCCGGCA	TTAATTTATC	7020
7021	AGCTAGAACG	GTTGAATATC	ATATTGATGG	TGATTGACT	GTCTCCGGCC	TTTCTCACCC	7080
7081	TTTTGAATCT	TTACCTACAC	ATTACTCAGG	CATTGCATTT	AAAATATATG	AGGGTTCTAA	7140
7141	AAATTTTTAT	CCTTGCGTTG	AAATAAAGGC	TTCTCCCGCA	AAAGTATTAC	AGGGTCATAA	7200
7201	TGTTTTGGT	ACAACCGATT	TAGCTTTATG	CTCTGAGGCT	TTATTGCTTA	ATTTTGCTAA	7260
7261	TTCTTTGCCT	TGCCTGTATG	ATTTATTGGA	CGTT			7294

FIG. 9-2

15/16

	10	20	30	40	50	60	
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAAT	60
61	ATAGCTAAAC	AGGTTATTGA	CCATTTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT	120
121	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA	180
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAGATTC	AGCAATTAAG	CTCTAAGCCA	240
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAGG	TACTCTCTAA	TCCTGACCTG	300
301	TTGGAGTTTG	CTTCCGGTCT	GGTTCGCTTT	CGAGCTCGAA	TTAAAACGCG	ATATTTGAAG	360
361	TCTTTGCGGC	TTCTCTTAA	TCTTTTGTAT	GCAATCCGCT	TTCTGTTCTGA	CTATAATAGT	420
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA	480
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCGCGAG	TATTGGACGC	TATCCAGTCT	540
541	AAACATTTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTTG	CAAAAGCCTC	TCGCTATTTT	600
601	GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCTCGT	660
661	AATTCCTTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCCTAA	ATCTCAACTG	720
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATTTT	780
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA	840
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGTIT	900
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG	960
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTG6TC	1020
1021	TGTACACCGT	TCATCTGTCC	TCTTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC	1080
1081	GTCTGCGCCT	CGTTCGGGCT	AAGTAACATG	GAGCAGGTCG	CGGATTTCTGA	CACAATTTAT	1140
1141	CAGGCGATGA	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT	1200
1201	CAAAGATGAG	TGTTTTAGTG	TATTCCTTCG	CCTCTTTCGT	TTTAGGTTGG	TGCCTTCGTA	1260
1261	GTGGCATTAC	GTATTTTACC	CGTTTAAATG	AACTTCCTC	ATGAAAAAGT	CTTTAGTCCT	1320
1321	CAAAAGCCTG	GTAGCCGTTG	CTACCCTCGT	TCCGATGCTG	TCTTTCGCTG	CTGAGGGTGA	1380
1381	CGATCCCGCA	AAAGCGGCCCT	TAACTCCCTT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440
1441	TGCGTGCGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAAGAA	1500
1501	ATTACACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAG	GGCTCCTTTT	GGAGCCTTTT	1560
1561	TTTTTGGAAGA	TTTCAACGT	GAAAAAATTA	TTATTCGCAA	TTCTTTTAGT	TGTTCTTTTC	1620
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA	1680
1681	TTTACTAACG	TCTGGAAGAA	GCACAAAAC	TTAGATCGTT	ACGCTAATA	TGAGGGTTGT	1740
1741	CTGTGGAATG	CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTTACA	1800
1801	TGGGTTCTTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGC6GT	1860
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920
1921	ATTCCGGGCT	ATACTTATAT	CAACCCTCTC	GACGGCACTT	ATCCGCGTGG	TACTGAGCAA	1980
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCTATGTT	2040
2041	CAGAATAATA	GGTTCGGAAG	TAGGCAGGGG	GCATTAACCTG	TTTATACGGG	CAGTGTACTT	2100
2101	CAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTAACTC	CTGTATCATC	AAAAGCCTATG	2160
2161	TATGACGCTT	ACTGGAACGG	TAAATTGAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA	2220
2221	GATCCATTCTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT	2280
2281	GCTGGCGGCG	GCTCTGGTGG	TGTTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
2341	GCGGTTCTTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTTCGGGT	2400
2401	GATTTTGATT	ATGAAAAGAT	GGCAACGCGT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460
2461	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATACCGGT	2520
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	2580
2581	GGTGATTTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCACCT	2640
2641	TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT	2700
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTGACAA	AATAAACTTA	2760
2761	TTCCGTGGTG	TCTTTGCGTT	TCTTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTTCTACG	2820
2821	TTTGCTAACA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTTG	GGTATTCCGT	2880
2881	TATTATTGCG	TTTCTCGGT	TTCTTCTGG	TAACTTTGTT	CGGCTATCTG	CTTACTTTTC	2940
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCAAT	GTTTCTTGCT	CTTATTATTG	3000
3001	GGCTTAACTC	AATCTTGTG	GGTTATCTCT	CTGATATTAG	CGCTCAATTA	CCCTCTGACT	3060
3061	TTGTTCAAGG	TGTTTCAGTTA	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTTT	TATGTTATTC	3120
3121	TCTCTGTAAA	GCGCTGATT	TTCATTTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTG6	3180
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAACCTGCA	AATTAGGCTC	TGGAAAGACG	3240
3241	CTCGTTAGCG	TTGGTAAGAT	TAGGATAAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT	3300
3301	CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTCCGGAGGT	TCGCTAAAAC	GCCTCGCGTT	3360
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGGCG	CGGTAATGAT	3420
3421	TCCTACGATG	AAAATAAAAA	CGGCTTGCTT	GTCTCGATG	AGTGCGGTAC	TGGGTTTAA	3480
3481	ACCCGTTCTT	GGAATGATAA	GGAAGACAGC	CCGATTATTG	ATTGGTTTCT	ACATGCTCGT	3540
3541	AAATTAGGAT	GGGATATTAT	TTTTCTTGTT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG	3600
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTTAT	TGTCGTCGTC	TGGACAGAAT	TACTTTACCT	3660
3661	TTTGTGCGTA	CTTTATATTC	TCTTATTACT	GGCTCGAAAA	TGCCTCTGCC	TAAATTACAT	3720
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TAAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT	3780
3781	ACTGGTAAGA	ATTTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT	3840

FIG. 10-1

SUBSTITUTE SHEET

16 / 16

3841	TCCGGTGT	ATTCTTAT	AACGCCTT	TTATCACAC	GTCGGTAT	CAAACCAT	3900
3901	AATTTAGG	AGAAGATG	GCTTACTA	ATATATTT	AAAAGTTT	ACGCGTTC	3960
3961	TGCTTGGC	TTGGATTG	ATCAGCAT	ACATATAG	ATATAACCC	ACCTAAGC	4020
4021	GAGGTTAA	AGGTAGTC	TCAGACCT	GATTTTGAT	AATTCACAT	TGACTCTTC	4080
4081	CAGCGTCT	ATCTAAGC	TCGCTATG	TTCAAGGAT	CTAAGGGAA	ATTAATTA	4140
4141	AGCGACGA	TACAGAAG	AGGTTATT	CTCACATAT	TTGATTTAT	TACTGTTTC	4200
4201	ATTAATAA	GTAATTCAA	TGAAATGTT	AAATGTAAT	AATTTTGTT	TCTTGATGT	4260
4261	TGTTTCAT	TCTTCTTT	CTCAGGTA	TGAAATGA	AATTCGCCT	TGCGCGATT	4320
4321	TGTAACCT	TATTCAAAG	AATCAGGCA	ATCCGTTAT	GTTTCTCCC	ATGTAAGAG	4380
4381	TACTGTTA	GTATATTC	CTGACGTT	ACCTGAAAT	CTACGCAAT	TCTTTATTC	4440
4441	TGTTTACG	GCTAATAAT	TTGATATGG	TGGTTCAAT	CCTTCCATA	TTCAGAAGT	4500
4501	TAATCCAA	AATCAGGAT	ATATTGATG	ATTGCCATC	TCTGATAAT	AGGAATATG	4560
4561	TGATAATTC	GCTCCTTCT	GTGGTTTCT	TGTTCCGCA	AATGATAAT	TTACTCAAAC	4620
4621	TTTTAAAT	AATAACGTT	GGGCAAGGA	TTTAATACG	GTTGTGCAAT	TGTTTGTAAG	4680
4681	GTCTAATA	TCTAAATCT	CAAAATGAT	ATCTATTGAC	GGCTCTAAT	TATTAGTTG	4740
4741	TAGTGCAC	AAAGATATT	TAGATAAC	TCCTCAATC	CTTCTACTG	TTGATTGGC	4800
4801	AACTGACC	ATATTGATT	AGGGTTTGT	ATTGAGGTT	CAGCAAGGT	ATGCTTTAG	4860
4861	TTTTTCAT	GCTGCTGG	CTCAGCGTG	CACGTTGCA	GGCGGTGTT	ATACTGACC	4920
4921	CCTCACCT	GTTTTATCT	CTGCTGGTG	TTCTGTTCC	ATTTTTAAT	GCGATGTTT	4980
4981	AGGGCTAT	GTTTCGCG	TAAAGACTA	TAGCCATT	AAAATATTG	CTGTGCCAC	5040
5041	TATCTTAC	CTTTCAGGT	AGAAGGTT	TATCTCTGT	GGCCAGAA	TCCCTTTAT	5100
5101	TACTGGTC	GTGACTGGT	AATCTGCCA	TGTAATAAT	CCATTTTCA	CGATTTGAG	5160
5161	TCAAAATG	GGTATTTCC	TGAGCGTTT	TCCTGTTGA	ATGGCTGGC	GTAATATTG	5220
5221	TCTGGATAT	ACCAGCAAG	CCGATAGTT	GAGTTCTTC	ACTCAGGCA	GTGATGTTT	5280
5281	TACTAATCA	AGAAGTATT	CTACAACGT	TAATTTGCG	GATGGACAG	CTCTTTTAC	5340
5341	CGGTGGCCT	ACTGATTAT	AAAACACTT	TCAAGATT	GGCGTACCG	TCCTGTCTA	5400
5401	AATCCCTTA	ATCGGCCCT	TGTTTAGCT	CCGCTCTGT	TCCAACGAG	AAAGCACGT	5460
5461	ATACGTGCT	GTCAAAGCA	CCATAGTAG	CGCCCTGTG	CGGCGCAT	AGCGCGCGG	5520
5521	GTGTGGTGG	TACGCGCAG	GTGACCGCT	CACTTGCCG	CGCCCTAGC	CCCGCTCCT	5580
5581	TCGCTTTCT	CCCTTCCCT	CTCGCCACG	TCGCGGCTT	TCCCGTCAA	GCTCTAAAT	5640
5641	GGGGGCTCC	TTTAGGGTT	CGATTTAGT	CTTTACGGC	CCTCGACCC	AAAAAACTG	5700
5701	ATTTGGGTG	TGGTTCACG	AGTGGGCGA	CGCCCTGAT	GACGGTTTT	CGCCCTTTG	5760
5761	CGTTGGAGT	CACGTTCTT	AATAGTGGC	TCTTGTCCA	AACTGGAAC	ACACTCAAC	5820
5821	CTATCTCGG	GATTTATAA	GATTTATAA	GGATTTTGC	GATTTGGA	CCACCATCA	5880
5881	ACAGGATTT	CTATTCTTT	GGCAAACCA	CGTGGACCG	TTGCTGCA	TCTCTCAGG	5940
5941	CCAGGCGGT	AGGGCAATC	AGCTGTTGC	CGTCTCGCT	GTGAAAAG	AAACCACCT	6000
6001	GGCGCCCA	ACGCAAACG	CCTCTCCCC	CGGTTGGCC	GATTCAATA	TGCAGCTGG	6060
6061	ACGACAGGT	TCCCGACTG	AAAGCGGCA	GTGAGCGCA	CGCAATTA	GTGAGTTAG	6120
6121	TCACTCATT	GGCACCCAG	GCTTTACAT	TTATGCTTC	GGCTCGTAT	TTGAGTTGA	6180
6181	TTGTGAGCG	ATAACAATT	CACACGCGT	ACTTGGCA	GGCCGTCGT	TTACAACGT	6240
6241	GTGACTGGG	AAACCTGGC	GTTACCCA	CTTTGTAC	GGAGAAATA	AAGTGAACA	6300
6301	AAGCACTAT	GCACTGGC	TCTTACCGT	ACTGTTTAC	CCTGTGGCA	AAGCCCTCT	6360
6361	GAGGCATCC	GGAGCTGA	GCGATGACC	TGCTAAGG	GCATTCAAT	GTTTACAGG	6420
6421	AAGTGCTACT	GAGTACATT	GCTACGCTT	GGCTATGTA	GATGTTATG	TTGGTGCTA	6480
6481	CATAGGGAT	AAATTATT	AAAAGTTT	GAGCAAGG	TCTTAAGCA	TAGCGAAG	6540
6541	GCCCGCACC	ATCGCCCTC	CCAACAGTT	CGCAGCCTG	ATGGCGAAT	GCGCTTTGC	6600
6601	TGGTTTCCG	CACCAGAAG	GGTGCCGGA	AGCTGGCTG	AGTGCGATC	TCCTGAGGC	6660
6661	GATACGGTC	TCGTCCCCT	AAACTGGC	ATGCGGTT	ACGATGCGC	CATCTACAC	6720
6721	AACGTAACT	ATCCCATT	GGTCAATCC	CCGTTTGTT	CCACGGAG	TCCGACGGT	6780
6781	TGTACTCGC	TCACATTTA	TGTTGATGA	AGCTGGCTA	AGGAAGGCC	GACGCGAAT	6840
6841	ATTTTTGAT	GCCTTCTAT	TGGTAAAA	ATGAGCTGT	TTAACAATA	TTTAACGCG	6900
6901	ATTTTAACA	AATATTACG	TTTACAATT	AAATATTGC	TTATACAAT	TTCTGTTT	6960
6961	TGGGGCTTT	CTGATTATC	ACCGGGGT	ATATGATTG	CATGCTAGT	TTACGATT	7020
7021	CGTTCATCG	TTCTCTTGT	TGCTCCAG	TCTCAGGCA	TGACCTGAT	GCCTTTGT	7080
7081	ATCTCTCAA	AATAGCTAC	CTCTCCGCA	TTAATTTAT	AGCTAGAAG	GTTGAATAT	7140
7141	ATATTGATG	TGATTGACT	AAAATATAT	TTTCTCACC	TTTGAATCT	TTACCTAC	7200
7201	ATTACTCAG	CATTGCAAT	AAAGTATT	AGGGTCTAA	AAATTTTAT	CCTTGCCTG	7260
7261	AAATAAAGG	TTCTCCGCA	AAAGTATT	AGGGTCATA	TGTTTTTGT	ACAACCGAT	7320
7321	TAGCTTTAT	CTCTGAGGT	TTATTGCTA	ATTTTGCTA	TTCTTTGCT	TGCTGTATG	7380
7381	ATTTATTGA	CGTT					7394

FIG. 10-2

SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/07141

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) ⁶		
According to International Patent Classification (IPC) or to both National Classification and IPC IPC(5): C12N 1/24; 15/00; C07H 21/00 U.S. CL.: 435/252.33, 320.1, 172.3; 536/27		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
U.S.	435/252.33, 320.1, 172.3, 69.1; 536/27	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁸		
APS, CAS: search terms: Codon bins, codon preference		
III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹		
Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
Y	EP. A. 0.383.620 (Cook) 22 August 1990. See entire document.	1-87
Y	US. A. 4,458,066 (Caruthers et al.) 03 July 1984. see entire document.	1-87
Y	US. A. 4,771,000 (Verrips et al.) 13 September 1988. see entire document.	8.9.24-26 32-34. 55-57. 64-66. 73-75. 81-87
Y	APPLIED MICROBIOLOGY AND BIOTECHNOLOGY. Volume 21. issued 1985. J.M. Jaynes et al., "Construction and expression of synthetic DNA fragments coding for polypeptides with elevated levels of essential amino acids". pages 200-205. see entire document.	1-87
<p>¹⁰ Special categories of cited documents: 10</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"Z" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report
11 December 1991		22 JAN 1992
International Searching Authority		Signature of Authorized Officer
ISA/US		James Ketter ebw

III DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	GENE, Volume 44, issued 1986. A.R. Oliphant. "Cloning of random-sequence oligodeoxynucleotides", pages 177-183, see entire document.	1-87
Y	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE, Volume 87, issued August 1990, Gwirla et al., "Peptides on phage: A vast library of peptides for identifying ligands", pages 6378-6382, see entire document.	1-87
Y	SCIENCE, Volume 249, issued 27 July 1990, J.J. Devlin, "Random Peptide Libraries: A Source of Specific Protein Binding Molecules", pages 404-406, see entire document.	1-87
Y	SCIENCE, Volume 249, issued 27 July 1990, J.K. Scott, "Searching for Peptide Ligands with an Epitope Library", pages 386-390, see entire document.	1-87
Y	E.-L. WINNACKER, "From Genes to Clones: Introduction to Gene Technology", published 1987 by VCH VmbH (Weinheim, Germany), See pages 276-279, especially Table 7-4.	1-87
Y	SCIENCE, Volume 228, issued 14 June 1985, G.P. Smith, "Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface", pages 1315-1317, see entire document.	8.9.24-26. 32-34. 55-57. 64-66. 73-75. 81-87